MANCHESTER 1824

The University of Manchester

Economics
Discussion Paper Series
EDP-1704

Bootstrapping Structural Change Tests

Otilia Boldea

Alastair R. Hall

Adriana Cornea-Madeira

February 2017

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

Bootstrapping Structural Change Tests

Otilia Boldea

Tilburg University

Adriana Cornea-Madeira

University of York

and

Alastair R. Hall

University of Manchester

January 10, 2017

Abstract

Bootstrap methods have been applied extensively in testing for structural breaks in the past few decades, but the conditions under which they are valid are, for the most part, unknown. In this paper, we fill this gap for the empirically important scenario in which supremum-type tests are used to test for discrete parameter change in linear models estimated by least squares methods. Our analysis covers models with exogenous regressors estimated by Ordinary Least Squares (OLS), and models with endogenous regressors estimated by Two Stage Least Squares (2SLS). Specifically, we show the asymptotic validity of the (IID and wild) recursive and fixed-regressors bootstraps for inference based on sup-F and sup-Wald statistics for testing both the null hypothesis of no parameter change versus an alternative of parameter change at $k > 0$ unknown break points, and also the null hypothesis of parameter change at ℓ break points versus an alternative of parameter change at $\ell + 1$ break points. For the case of exogenous regressors, Bai and Perron (1998) derive and tabulate the limiting distributions of the test statistics based on OLS under the appropriate null hypothesis; for the case of endogenous regressors, Hall, Han, and Boldea (2012) show that the same limiting distributions hold for the analogous test statistics based on 2SLS when the first stage model is stable. As part of our analysis, we derive the limiting distribution of the test statistics based on 2SLS when the regressors are endogenous and the first stage regression exhibits discrete parameter change. We show that the asymptotic distributions of the second-stage break-point tests are non-pivotal, and as a consequence the usual Bai and Perron (1998) critical values cannot be used. Thus, our bootstrap-based methods represent the most practically feasible approach to testing for multiple discrete parameter changes in the empirically relevant scenario of endogenous regressors and an unstable first stage regression. Our simulation results show very good finite sample properties with all the versions of the bootstrap considered here, and indicate that the bootstrap tests are preferred over the asymptotic tests, especially in the presence of conditional heteroskedasticity of unknown form.

JEL classification: C12, C13, C15, C22

Keywords: Multiple Break Points; Instrumental Variables Estimation; Two-stage Least Squares; IID bootstrap; Wild bootstrap; Recursive bootstrap; Fixed-regressor bootstrap; Conditional heteroskedasticity.

The first author acknowledges the support of the VENI Grant 451-11-001. The second author acknowledges the support of the British Academy's PDF/2009/370. The third author acknowledges the support of ESRC grant RES-062-23-1351. We thank Xu Cheng, Russell Davidson, Frank Diebold, Sílvia Gonçalves 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 for useful comments.

1 Introduction

Economists routinely apply structural change tests. Structural changes (breaks) affect models for the evolution of key economic and financial time series such as GDP, inflation, exchange rates, interest rates, stock returns, money demand and income.[1] Structural breaks 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. A rich literature on asymptotic tests used to detect structural breaks has developed in recent years, see among others Andrews (1993), Andrews and Ploberger (1994), Bai and Perron (1998), and Hall, Han, and Boldea (2012). However, given the current sample size of macroeconomic data, the asymptotic tests often provide inaccurate conclusions about the number and location of breaks. Therefore, different bootstrap methods may be preferred, as shown in Christiano (1992), Diebold and Chen (1996), Hansen (2000), Banerjee, Lazarova, and Urga (2002), de Peretti and Urga (2004), O'Reilly and Whelan (2004), Clark (2006), Levin and Piger (2006), Antoshin, Berg, and Souto (2008), Berg, Ostry, and Zettelmeyer (2012), Bergamelli and Urga (2013). With the exception of Hansen (2000)'s fixed-regressor bootstrap, the asymptotic validity of these different bootstrap methods for structural change tests has not been established in any of the above studies or elsewhere in the literature.

In this paper, we fill this gap for the empirically important scenario in which Supremum-type tests are used to test for discrete parameter change in linear models estimated by least squares methods. Within this context, two hypotheses are naturally of interest: (a) a null hypothesis of no parameter change versus an alternative of parameter change at a fixed number of break points with unknown location; (b) the null hypothesis of parameter change at ℓ break points against the alternative of parameter change at an additional break point of unknown location. These tests are routinely used in a sequential procedure to estimate the number of breaks, and the resulting estimator of the number of breaks approaches the true number of breaks with probability one in the limit, provided that the significance level of each test shrinks to zero slow enough.[2] Our analysis covers inferences about the hypotheses in (a) and (b) based on F-type and Wald statistics in both models with exogenous regressors estimated by Ordinary Least Squares (OLS), and also models with endogenous regressors estimated by Two Stage Least Squares (2SLS). We consider the nonparametric *IID bootstrap* (which treats the regression errors as independent and identically distributed (IID) from an unspecified distribution) and the *wild bootstrap* (which allows for the regression errors to have unconditional hetereskedasticity and conditional heteroskedasticity in the form of (G)ARCH and stochastic volatility models).[3] Within this framework, we show the asymptotic validity of the IID and wild *recursive bootstrap* (thereafter IR and WR bootstrap), which generates recursively the bootstrap observations, and of the IID and wild *fixed-regressor bootstrap* (thereafter IF and WF bootstrap), which keeps all the (lagged) regressors fixed, i.e. it simply adds the (IID or wild) bootstrap residuals to the estimated conditional mean.[4] Our analysis, therefore, proves the first order validity of the bootstrap applied in the papers cited in the previous paragraph.[5] To our knowledge, the form of the fixed bootstrap we consider is new, although it is a

[1] See for example: Bergamelli and Urga (2013); Christiano (1992); de Peretti and Urga (2004); Feldstein and Stock (1994); Hall, Han, and Boldea (2012); Hansen (2001); Hsu (2005); Levin and Piger (2006); Morana and Beltratti (2004); Perron (1989); Perron and Vogelsang (1992); Quintos, Fan, and Phillips (2001); Stock and Watson (2005); Stock (1994).

[2] See Bai and Perron (1998), Proposition 8.

[3] As illustrated in Bollerslev (1986), Hodrick (1992), Bekaert and Hodrick (2001), and Gonçalves and Kilian (2004), many financial and macroeconomic series exhibit conditional heteroskedasticity.

[4] As opposed to the wild recursive bootstrap, the wild fixed-regressor bootstrap allows for a richer structure of conditional heteroskedasticy such as (G)ARCH and stochastic volatility models with asymmetric errors.

[5] More exactly, Christiano (1992) employs the IR bootstrap to test for a break in the U.S. GNP. Diebold and Chen (1996) provide

natural alternative to the recursive bootstrap.

Our proofs rest on showing that the bootstrap version of the tests has the same limiting distribution as the analogous statistic in the sample. For the case of exogenous regressors, Bai and Perron (1998) derive and tabulate the limiting distributions of the test statistics above based on OLS under the appropriate null hypothesis; for the case of endogenous regressors, Hall, Han, and Boldea (2012) show that the same limiting distributions hold for the analogous test statistics based on 2SLS when the first stage model is stable. As part of our analysis, we derive the limiting distribution of the test statistics based on 2SLS when the regressors are endogenous and the first stage regression exhibits discrete parameter change. It is shown that the asymptotic distributions of the second-stage break-point tests are non-pivotal, and as a consequence the usual Bai and Perron (1998) critical values can not be used. Our simulation results indicate that the differences between our limiting distributions and the Bai and Perron (1998) distributions are non-trivial. This feature of the tests was recognized (although not proved) by Hall, Han, and Boldea (2012) who proposed to divide the sample into subsamples within which the first-stage equation is stable, and to test for parameter change in these sub-samples using the sup-F and sup-$Wald$ asymptotic tests for a stable first stage. However, these subsamples may be small given the current size of macroeconomic datasets, meaning that it is infeasible to calculate the required tests or even if feasible, the tests based on the subsamples are unreliable. In contrast, we propose bootstrapping the sup-F and sup-$Wald$ tests that are calculated from the full sample and that take account of the breaks in the first stage. Thus, our bootstrap-based methods may represent the only practically feasible approach to testing for multiple discrete parameter change in the empirically relevant scenario of endogenous regressors and an unstable first stage regression with macroeconomic data.

It is not our aim here to establish that the bootstrap provides a superior approximation to the conventional asymptotic tests in Bai and Perron (1998), in Hall, Han, and Boldea (2012) and in this paper. Instead, we provide the assumptions under which both the asymptotic and bootstrap tests are asymptotically valid. Nevertheless, since the bootstrap incorporates sample information, we expect that the bootstrap provides a superior approximation to the asymptotic tests which do not incorporate similar sample information. This expectation is confirmed by our simulation results provided at the end of this paper. In addition, our simulations show that in the presence of conditional heteroskedasticity, the WR and WF bootstraps outperform the asymptotic tests in samples as large as 480, even though the latter tests provide asymptotically valid inference.

There are several differences between the bootstraps in this paper and earlier work on the bootstrap. First, our fixed-regressor bootstrap is applied as in Kreiss (1997) and Gonçalves and Kilian (2004) in that a bootstrap sample is obtained by adding the (IID or wild) bootstrap residuals to the conditional mean, as opposed to Hansen (2000)'s fixed-regressor bootstrap which simply takes the (IID or wild) bootstrap residuals as the bootstrap sample

simulation results for testing for a single break in stationary autoregressive models using the IR bootstrap and conclude that the bootstrap is preferred to the asymptotic test proposed by Andrews (1993). Clark (2006) confirms the conclusions of Diebold and Chen (1996) by considering a larger diversity of data generating processes drawn from 1984-2002 estimates of autoregressive models used for modeling inflation. In addition, Antoshin, Berg, and Souto (2008) show in Monte Carlo simulations that the IR bootstrap is preferred to the asymptotic tests for multiple breaks proposed by Bai and Perron (1998), and Berg, Ostry, and Zettelmeyer (2012) use the IR bootstrap to show evidence of structural breaks in the economic growth of 140 countries. Moreover, O'Reilly and Whelan (2004) show by simulations that the WR bootstrap works well across a wide variety of data generating processes used in macroeconomics. Building on this conclusion, Levin and Piger (2006) use the WR bootstrap to find evidence of structural breaks in the intercept of autoregressive models for inflation for eight OECD countries. Finally, Banerjee, Lazarova, and Urga (2002), de Peretti and Urga (2004) and Bergamelli and Urga (2013) show through Monte Carlo simulations that the IR bootstrap detects multiple breaks in systems of equations.

for the dependent variable. Second, our analysis allows for shrinking breaks in the marginal distribution of the regressors and in the unconditional variance of the regression errors without requiring that the econometrician estimates these breaks, whereas Hansen (2000) fixed-regressor bootstrap allows for the possibility of fixed breaks in the marginal distribution of the regressors without having to estimate the breaks. Third, while Gonçalves and Kilian (2004) focus on the WR and WF bootstrap in the presence of conditional heteroskedasticity only, we allow for both conditional heteroskedasticity and unconditional heteroskedasticity provided a global homoskedasticity assumption is satisfied, similar to Cavaliere, Rahbek, and Taylor (2010) in the context of cointegration rank testing.[6] Finally, our bootstraps for structural change tests rely on the existence of moments slightly larger than four, as opposed to Gonçalves and Kilian (2004) who assume the existence of moments of order eight in the context of bootstrapping the coefficients of autoregressive models.[7]

The paper is organised as follows. In Section 2 we introduce the model with possibly endogenous regressors and the multiple break point tests. In Section 3 we introduce the assumptions needed to derive the asymptotic distribution of the multiple break point tests. In Section 4 we derive the asymptotic distribution of the multiple break point tests when the model has endogenous regressors and (un)stable first-stage equations, while in Section 5 we introduce the multiple break point bootstrap tests, their assumptions, and show the asymptotic validity of the IR, WR, IF, and WF bootstraps when the first-stage is (un)stable, or when the regressors are exogenous. We illustrate the finite sample performance of our bootstrap tests in a simulation experiment in Section 6 and compare it with the finite sample performance of the asymptotic tests. Section 7 concludes. Appendix A contains the VARX and VMAX representations of the model in (1) and (3). Appendix B contains definitions, and Appendix C tables for the simulation section. All the proofs are relegated to the Supplemental Appendix.

Notation. The symbol $[\cdot]$ denotes the integer part; $\text{rk}(\cdot)$ denotes the rank of a matrix; bold small letters (latin or greek) denote in general vectors, i.e. $\boldsymbol{b}$; $\boldsymbol{W}(\cdot)$ denotes a vector of standard independent Brownian motions; bold capital letters (latin or greek) denote matrices, i.e. $\boldsymbol{\Sigma}$; $\boldsymbol{I}_a$ is the $a \times a$ identity matrix; $\otimes$ is the Kronecker product; $\mathbf{0}_a$ denotes a $a \times 1$ vector of zeros; $\mathbf{O}_{a\times a}$ denotes a $a \times a$ matrix of zeros; $\dim(\boldsymbol{b})$ denotes the number of rows of the vector $\boldsymbol{b}$; $\text{diag}(a_1, \ldots, a_n)$ denotes a diagonal matrix with element $a_1, \ldots, a_n$ on the main diagonal; $1_{\mathcal{J}}$ is the indicator function, and equals 1 if condition $\mathcal{J}$ is satisfied, and 0 otherwise; $\Rightarrow$ denotes weak convergence in Skorohod metric; E, var, cov denote the expectation, variance and covariance under the probability measure P of the data; $\xrightarrow{p}$ denotes convergence in probability under P; for a scalar random variable x, the L_p-norm of x is $||\ x\ ||_p = (\text{E}\mid x\mid^p)^{1/p}$, $p > 0$, where $|\cdot|$ denotes the absolute value; $|\boldsymbol{A}|$ denotes the determinant of a square matrix $\boldsymbol{A}$; 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 data; $\vee$ denotes the maximum and $\wedge$ denotes the minimum. Corresponding to the Euclidian vector norm $\|\boldsymbol{v}\| = \left(\sum_{i=1}^p v_i^2\right)^{1/2}$, we define the matrix norm $\|\boldsymbol{A}\| = \sup_{\boldsymbol{v}\neq\mathbf{0}} \|\boldsymbol{A}\boldsymbol{v}\| / \|\boldsymbol{v}\|$ for generic matrix $\boldsymbol{A}$ of size $q \times p$ and vector $\boldsymbol{v}$ of size $p \times 1$. Finally, $U(T_a, T_b)$ denotes the uniform distribution on the interval $[T_a, T_b]$, which generates integer values between T_a and T_b, and $IID(0, 1)$ denotes independently and identically distributed with mean zero and variance one.

[6] For example, this allows for models with seasonal heteroskedasticity.

[7] The difference stems from the fact that Gonçalves and Kilian (2004) use a Functional Central Limit Theorem (FCLT) that requires the convergence of the sample moments while we use the FCLT of Wooldridge and White (1988), Theorem 2.11 which requires only verifying the convergence of the population moments.

2 Model and test statistics

This section introduces the dynamic linear model and the multiple break tests calculated over the full sample, in the general case with a mix of endogenous and exogenous regressors and with an unstable first-stage equation. By construction, this framework can be specialized to cover the other two cases discussed in the introduction: if all the regressors are exogenous, then the endogenous regressors can be dropped from the analysis; if some regressors are endogenous but the first stage is stable, then the analysis is simplified to no breaks in the first stage. Both these special cases are discussed further below. In Section 3 we present the assumptions necessary for the derivations of the asymptotic distributions of the multiple break tests which are given in Section 4. However, note that we adopt the so-called "shrinking breaks" assumption for both the equation of interest and the first stage regression. Under this assumption, the parameters are different across different regimes for finite samples but are converging toward some common limiting value at a controlled rate. This assumption is common in the literature on structural break literature, *e.g.* Bai (1997), Bai and Perron (1998) (BP henceforth), and is designed to provide an asymptotic theory that approximates the finite sample behaviour of statistics of interest when the breaks are of "moderate" size.

To facilitate presentation, we first define the notion of *partitions*. For the equation of interest, if there are m breaks in the sample $1,\dots,T$, at $T_1,\dots,T_m$, then an m-partition (of the sample) is defined through the break fractions, and we write: $\boldsymbol{\lambda} = (0,\lambda_1,\dots,\lambda_m,1)$, where λ_i are the break fractions, i.e. $T_i = [T\lambda_i]$, for $i = 0,1,\dots,m+1$, and $T_0 = 0$, $T_{m+1} = T$. For the first-stage equation, if there are h breaks in the sample $1,\dots,T$, at values $T_1,\dots,T_h$, then the h-partition (of the sample) is defined through the break fractions: $\boldsymbol{\pi} = (0,\pi_1,\dots,\pi_h,1)$, where $T_i = [T\pi_i]$, for $i = 0,1,\dots,m+1$, and $T_0 = 0$, $T_{m+1} = T$. The true partitions are denoted $\boldsymbol{\lambda}^0$, $\boldsymbol{\pi}^0$, with break points $T_i^0 = [T\lambda_i^0]$, $T_j^* = [T\pi_j^0]$ and break fractions λ_i^0, π_j^0, and the estimated partitions are denoted by $\hat{\boldsymbol{\lambda}}$, $\hat{\boldsymbol{\pi}}$, with estimated break points: $\hat{T}_i = [T\hat{\lambda}_i]$, $\hat{T}_j^* = [T\hat{\pi}_j]$ or estimated break fractions $\hat{\lambda}_i, \hat{\pi}_j$, for $i = 1,\dots,m$ and $j = 1,\dots,h$.

As in Hall, Han, and Boldea (2012) (HHB henceforth), the equation of interest is a linear regression model with m breaks ($m+1$ regimes), that is:

$$y_t = \underbrace{\boldsymbol{x}_t'}_{1\times p_1}\underbrace{\boldsymbol{\beta}_{\boldsymbol{x},(i)}^0}_{p_1\times 1} + \underbrace{\boldsymbol{z}_{1,t}'}_{1\times p_2}\underbrace{\boldsymbol{\beta}_{\boldsymbol{z}_1,(i)}^0}_{p_2\times 1} + u_t = \underbrace{\boldsymbol{w}_t'}_{1\times p}\underbrace{\beta_{(i)}^0}_{p\times 1} + u_t, \quad i = 1,\dots,m+1, \quad t = T_{i-1}^0+1,\dots,T_i^0, \quad (1)$$

where $p = p_1 + p_2$, $\boldsymbol{w}_t = (\boldsymbol{x}_t', \boldsymbol{z}_{1,t}')'$, $\boldsymbol{\beta}_{(i)}^0 = \left(\boldsymbol{\beta}_{\boldsymbol{x},(i)}^{0'}, \boldsymbol{\beta}_{\boldsymbol{z}_1,(i)}^{0'}\right)'$, $\boldsymbol{\beta}_{\boldsymbol{z}_1,(i)}^0$ is the coefficient on the exogenous regressors $\boldsymbol{z}_{1,t}$ for subsample i and $\boldsymbol{\beta}_{\boldsymbol{x},(i)}^0$ is the coefficient on the endogenous regressors $\boldsymbol{x}_t$ for subsample i, and u_t is a mean zero disturbance correlated with $\boldsymbol{x}_t$. The exogenous regressors include lags of y_t and distributed lags of $\boldsymbol{x}_t$:

$$\underbrace{\boldsymbol{z}_{1,t}'}_{1\times p_2} = (\underbrace{\boldsymbol{r}_{1,t}'}_{1\times q_1}, \underbrace{y_{t-1}}_{1\times 1}, \underbrace{\boldsymbol{x}_{t-1}'}_{1\times p_1}, \underbrace{y_{t-2}}_{1\times 1}, \underbrace{\boldsymbol{x}_{t-2}'}_{1\times p_1}, \dots, \underbrace{\boldsymbol{x}_{t-\tilde{q}_1}'}_{1\times p_1}), \quad (2)$$

where $\boldsymbol{r}_{1,t}$ includes the intercept and other exogenous regressors (possibly lagged) and $p_2 = q_1 + \tilde{p}_1 + \tilde{q}_1 p_1$, where $\tilde{p}_1$ are the number of y_t lags included in $\boldsymbol{z}_{1,t}$ and $\tilde{q}_1$ are the number of $\boldsymbol{x}_t$ lags included in $\boldsymbol{z}_{1,t}$. Given that $\boldsymbol{x}_t$ is endogenous, it is plausible that (1) belongs to a system of structural equations and thus, for simplicity, we refer to (1) as the "structural equation" (SE). The reduced form equation (RF) for the endogenous regressors $\boldsymbol{x}_t$ is a linear regression model with h breaks ($h+1$ regimes), that is:

$$\underbrace{\boldsymbol{x}_t'}_{1\times p_1} = \underbrace{\boldsymbol{z}_t'}_{1\times q}\underbrace{\boldsymbol{\Delta}_{(i)}^0}_{q\times p_1} + \underbrace{\boldsymbol{v}_t'}_{1\times p_1}, \qquad i = 1,\dots,h+1, \qquad t = T_{i-1}^*+1,\dots,T_i^*, \quad (3)$$

where $\boldsymbol{\Delta}^0_{(i)}$ are the coefficients on the instruments $\boldsymbol{z}_t$ for subsample i, and h is treated as a known fixed number. To simplify the notation and the proofs, we assume throughout that when there are breaks in SE/ RF, the breaks occur in all the parameters, but the results in this paper generalize to the situation when only some parameters break. The instruments $\boldsymbol{z}_t$ include $\boldsymbol{z}_{1,t}$ and are uncorrelated with u_t and $\boldsymbol{v}_t$. They are defined as:

$$\begin{aligned}\underbrace{\boldsymbol{z}_t'}_{1\times q} &= (\underbrace{\boldsymbol{r}_{1,t}'}_{1\times q_1}, \underbrace{\boldsymbol{r}_{2,t}'}_{1\times q_2}, \underbrace{y_{t-1}}_{1\times 1}, \underbrace{\boldsymbol{x}_{t-1}'}_{1\times p_1}, \underbrace{y_{t-2}}_{1\times 1}, \underbrace{\boldsymbol{x}_{t-2}'}_{1\times p_1}, \ldots, \underbrace{y_{t-\tilde{p}}}_{1\times 1}, \underbrace{\boldsymbol{x}_{t-\tilde{p}}'}_{1\times p_1}) \\ &= (\underbrace{\boldsymbol{r}_t'}_{1\times(q_1+q_2)}, \underbrace{\tilde{\boldsymbol{y}}_{t-1}'}_{1\times(p_1+1)}, \underbrace{\tilde{\boldsymbol{y}}_{t-2}'}_{1\times(p_1+1)}, \ldots, \underbrace{\tilde{\boldsymbol{y}}_{t-\tilde{p}}'}_{1\times(p_1+1)}) = (\underbrace{\boldsymbol{r}_t'}_{1\times(q_1+q_2)}, \underbrace{\tilde{\tilde{\boldsymbol{y}}}_t'}_{1\times\tilde{p}(p_1+1)}). \end{aligned} \tag{4}$$

Here, $\tilde{\boldsymbol{y}}_{t-j}$, $j = 1, \ldots, \tilde{p}$, and $\tilde{\tilde{\boldsymbol{y}}}_t$ are defined as:

$$\tilde{\boldsymbol{y}}_{t-j}' = (y_{t-j}, \boldsymbol{x}_{t-j}'), \quad \tilde{\tilde{\boldsymbol{y}}}_t' = (\tilde{\boldsymbol{y}}_{t-1}', \ldots, \tilde{\boldsymbol{y}}_{t-\tilde{p}}') \tag{5}$$

and $\tilde{p} = \max(\tilde{p}_2, \tilde{q}_2)$, where $\tilde{p}_2$ is the number of y_t lags in $\boldsymbol{z}_t$ that have non-zero coefficients in (3), and $\tilde{q}_2$ are the number of $\boldsymbol{x}_t$ lags that have non-zero coefficients in (3). Note that because $\boldsymbol{z}_t$ includes $\boldsymbol{z}_{1,t}$, $\tilde{p} \geq \max(\tilde{p}_1, \tilde{q}_1)$.

To test for breaks in SE, we will explicitly consider only cases in which $m = 0, 1, 2$, but the results extend to $m > 2$ in the obvious way. Within this framework, two natural hypotheses of interest are:

(a) $H_0 : m = 0$ versus $H_1 : m = k$, with $k = 1, 2$;

(b) $H_0 : m = 1$ versus $H_1 : m = 2$.

To implement tests for (a) and (b), we need to consistently estimate the RF, and therefore, in general, we need to know the number of RF breaks h. We can estimate h via a BP sequential procedure in the RF: test the RF for zero versus one breaks, one versus two breaks, and so on, until one can no longer reject; the last null hypothesis gives the estimated number of RF breaks $\hat{h}$. Even though this is a sequential testing procedure, BP show that $\hat{h}$ approaches h with probability one as the sample size T grows, as long as the significance level in each step shrinks to zero slowly enough (see their Proposition 8). The same consistency result holds if we estimate h via the information criteria in Hall, Osborn, and Sakkas (2013). For this reason, in the rest of the paper, we treat h as known.

Our goal is to construct tests for hypotheses (a) and (b). We first estimate the unknown regression coefficients $\boldsymbol{\beta}^0_{(1)}, \ldots, \boldsymbol{\beta}^0_{(m+1)}$ and the m-break fraction partition $\boldsymbol{\lambda}^0$, via 2SLS.[8,9] In the first stage, the RF for $\boldsymbol{x}_t$ given in (3) is estimated using the BP OLS estimated h-partitions $\hat{\boldsymbol{\pi}}$, obtained equation by equation and pooled, substituted for the true h-partition $\boldsymbol{\pi}^0$.[10] Let $\hat{\boldsymbol{x}}_t$ denote the resulting predicted value for $\boldsymbol{x}_t$ from (3), which takes into account that there are h RF breaks. In the special case of stable RF, $h = 0$ and $\hat{\boldsymbol{x}}_t$ is estimated over the full sample. In the special case of no endogenous regressors, there is no $\boldsymbol{x}_t$, so estimation of the RF is skipped, and the analysis below is done without $\boldsymbol{x}_t$ and $\hat{\boldsymbol{x}}_t$, and with $\boldsymbol{z}_{1,t} = \boldsymbol{z}_t$.

In the general case, we let $\hat{\boldsymbol{w}}_t = (\hat{\boldsymbol{x}}_t', \boldsymbol{z}_{1,t}')'$, and in the second stage, we first estimate, for each m-partition $\boldsymbol{\lambda}$,

$$y_t \;=\; \hat{\boldsymbol{w}}_t'\boldsymbol{\beta}^0_{(i)} \;+\; \text{residuals}, \quad i = 1, ..., m+1; \quad t = T_{i-1}+1, ..., T_i. \tag{6}$$

[8] See the definition of partitions at the beginning of Section 2.

[9] Our preferred estimation method is the 2SLS since HHB show that the minimization of the GMM criterion yields inconsistent estimators of the break fractions, but the minimization of the 2SLS criterion yields consistent estimators of the break fractions.

[10] We can also use the multivariate methods in Qu and Perron (2007). For our purposes, it only matters that $\hat{\pi}_j - \pi^0_j = O_p(T^{2\rho-1})$, where ρ is defined in Assumption 5, and not how they are obtained. The fact that $\hat{\pi}_j - \pi^0_j = O_p(T^{2\rho-1})$ follows by Assumptions 4-8, 10, introduced in Section 3, and Theorem 2 in HHB; see also the discussion after Assumption 10 is introduced.

Therefore, the resulting OLS estimates $\hat{\boldsymbol{\beta}}_{(i)}(\boldsymbol{\lambda}) = (\hat{\boldsymbol{\beta}}'_{\boldsymbol{x},(i)}(\boldsymbol{\lambda}), \hat{\boldsymbol{\beta}}'_{\boldsymbol{z}_1,(i)}(\boldsymbol{\lambda}))'$ minimize:

$$SSR_m(\boldsymbol{\lambda};\, \boldsymbol{\beta}) = T^{-1} \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} \left(y_t - \hat{\boldsymbol{w}}_t' \boldsymbol{\beta}_{(i)} \right)^2 \tag{7}$$

with respect to $\boldsymbol{\beta} = \left(\boldsymbol{\beta}'_{(1)}, \boldsymbol{\beta}'_{(2)}, \ldots, \boldsymbol{\beta}'_{(m+1)} \right)'$. The break fraction estimates $\hat{\boldsymbol{\lambda}}$ are defined through:

$$\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{\operatorname{argmin}}\; SSR_m \left(\boldsymbol{\lambda};\, \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) \right), \tag{8}$$

where the minimization is taken over all possible partitions $\boldsymbol{\lambda}$. The corresponding break-point estimates are $\hat{T}_i = [T\hat{\lambda}_i]$. The 2SLS parameter estimates are $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\lambda}}) = \left(\hat{\boldsymbol{\beta}}'_{(1)}, \hat{\boldsymbol{\beta}}'_{(2)}, \ldots, \hat{\boldsymbol{\beta}}'_{(m+1)} \right)'$ associated with the estimated m-partition $\hat{\boldsymbol{\lambda}}$. Theorem 2 of HHB and Lemma 1.2 of the Supplemental Appendix of this paper show that, under the assumptions outlined in the next section, the break fractions in the RF and SE are converging fast enough so that the asymptotic distribution of the parameter estimators (which is key to deriving the asymptotic distributions of our structural change tests) is unaffected by the randomness in the break fraction estimates.

The hypotheses (a) and (b) outlined above are useful per se, but they can also be used to test the number of SE breaks m sequentially; this estimate of the number of SE breaks will also approach m with probability one in the limit, as in BP. We now consider hypotheses (a) and (b) in turn.

(a) The null hypothesis $H_0 : m = 0$ against the alternative hypothesis $H_1 : m = k$.

Denote

$$F_T(\boldsymbol{\lambda}) \;=\; \left(\frac{T - (k+1)p}{kp} \right) \left(\frac{SSR_0 - SSR_k(\boldsymbol{\lambda};\, \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}))}{SSR_k(\boldsymbol{\lambda};\, \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}))} \right) \tag{9}$$

where SSR_0 and SSR_k are the 2SLS sum of squared residuals, based on the fitted value for $\boldsymbol{x}_t$ under null and alternative hypothesis, using the k-partition $\boldsymbol{\lambda}$.[11] Define $\boldsymbol{\Lambda}_\epsilon = \{\boldsymbol{\lambda} : |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1 - \epsilon\}$. Then the sup-$F$ test statistic is defined as:

$$\sup\text{-}F_T \;=\; \sup_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}_\epsilon} F_T(\boldsymbol{\lambda}). \tag{10}$$

Note that to perform this test, it is necessary to pre-estimate the RF breaks if $h \geq 1$ and use them to compute the first stage projections; for that reason it is new compared to HHB, who only provide tests for a stable RF. The same statement holds for the sup-$Wald$ test defined below.

To write the sup-$Wald$ test of $H_0 : m = 0$ versus $H_1 : m = k$, we restate the null and alternative hypotheses in terms of linear restrictions on the parameters. Accordingly, 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 $j \neq i$, $j \neq i+1$. With this notation, the null and alternative can be equivalently stated as: $H_0 : \boldsymbol{R}_k \boldsymbol{\beta}^0 = 0$ versus $H_1 : \boldsymbol{R}_k \boldsymbol{\beta}^0 \neq 0$ where $\boldsymbol{\beta}^0 = (\boldsymbol{\beta}^{0'}_{(1)}, \ldots, \boldsymbol{\beta}^{0'}_{(k+1)})'$. The test statistic is:

$$\sup\text{-}Wald_T \;=\; \sup_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}_\epsilon} Wald_T(\boldsymbol{\lambda}), \tag{11}$$

$$Wald_T(\boldsymbol{\lambda}) \;=\; T\, \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})' \boldsymbol{R}_k' \left(\boldsymbol{R}_k \hat{\boldsymbol{V}}(\boldsymbol{\lambda}) \boldsymbol{R}_k' \right)^{-1} \boldsymbol{R}_k\, \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}), \tag{12}$$

[11] Note that if we compute SSR_0 with $\boldsymbol{w}_t$ instead of $\hat{\boldsymbol{w}}_t$, $SSR_0 = T^{-1} \sum_{t=1}^{T} \left(y_t - \boldsymbol{w}_t' \hat{\boldsymbol{\beta}} \right)^2$, and $SSR_0 - SSR_1$ (assuming $k = 1$) contains the term $T^{-1/2} \sum_{t=T_{i-1}+1}^{T_i} u_t \boldsymbol{x}_t$ which will explode since this term is not mean zero under the assumption that $\boldsymbol{x}_t$ is endogenous.

where:

$$\hat{\boldsymbol{V}}(\boldsymbol{\lambda}) = \text{diag}\left(\hat{\boldsymbol{V}}_{(1)}, \ldots, \hat{\boldsymbol{V}}_{(k+1)}\right), \qquad \hat{\boldsymbol{V}}_{(i)} = \hat{\boldsymbol{Q}}_{(i)}^{-1}\, \hat{\boldsymbol{M}}_{(i)}\, \hat{\boldsymbol{Q}}_{(i)}^{-1}, \qquad \hat{\boldsymbol{Q}}_{(i)} = T^{-1} \sum_{t=T_{i-1}+1}^{T_i} \hat{\boldsymbol{w}}_t \hat{\boldsymbol{w}}_t', \tag{13}$$

$$\hat{\boldsymbol{M}}_{(i)} \xrightarrow{p} \lim_{T\to\infty} \text{var}\left(T^{-1/2} \sum_{t=T_{i-1}+1}^{T_i} \boldsymbol{\Upsilon}^{0'}(t,T)\boldsymbol{z}_t\left(u_t + \boldsymbol{v}_t'\boldsymbol{\beta}^0_{\boldsymbol{x},(i)}\right)\right), \tag{14}$$

where $\boldsymbol{\Upsilon}^{0'}(t,T) = \sum_{j=1}^{h+1} 1_{t\in[T^*_{j-1}+1, T^*_j]}\boldsymbol{\Upsilon}^0_{(j)}$, with $\boldsymbol{\Upsilon}^0_{(j)}$ defined in Assumption 10 below.

(b) The null hypothesis $H_0 : m = 1$ against the alternative hypothesis $H_1 : m = 2$.
Following BP, a suitable statistic can be constructed as follows. For the model under the null hypothesis with one break, the estimated break point, denoted by $\hat{T}_1 = [T\hat{\lambda}_1]$, is obtained by a global minimization of the sum of the squared residuals as in (8). For the model under the alternative hypothesis with two breaks, one of the breaks is fixed at $\hat{T}_1$, and the location of the second break is chosen by minimizing the residual sum of squares in the sub-sample before $\hat{T}_1$ and in the sub-sample after $\hat{T}_1$. Formally, let $\boldsymbol{\Lambda}_1 = \{\tau_1 : \tau_1 = [\hat{T}_1\eta_1],\ \eta_1 \in [\epsilon, 1-\epsilon]\}$ and $\boldsymbol{\Lambda}_2 = \{\tau_2 : \tau_2 = \hat{T}_1 + (T - \hat{T}_1)\eta_2,\ \eta_2 \in [\epsilon, 1-\epsilon]\}$ be the sets of candidate break-points for the sub-sample before $\hat{T}_1$ and after $\hat{T}_1$, and

$$SSR_1(\hat{\tau}_1|\hat{T}_1) = \inf_{\tau_1\in\boldsymbol{\Lambda}_1} SSR_1(\tau_1|\hat{T}_1) \qquad SSR_2(\hat{\tau}_2|\hat{T}_1) = \inf_{\tau_2\in\boldsymbol{\Lambda}_2} SSR_2(\tau_2|\hat{T}_1), \tag{15}$$

with $\hat{\tau}_1$, $\hat{\tau}_2$ the arguments at which the two infima are obtained, $SSR_1(\tau_1|\hat{T}_1)$ the 2SLS sum of squared residuals in sample $[1, \hat{T}_1]$ with one break at $\tau_1 < \hat{T}_1$, and $SSR_2(\tau_2|\hat{T}_1)$ the 2SLS sum of squared residuals in sample $[\hat{T}_1 + 1, T]$ with one break at $\tau_2 > \hat{T}_1$. Then the sup-F statistic for testing for one break in SE against two breaks in SE is:

$$\sup\text{-}F_T(2|1) = \max\left(\frac{SSR_1(\hat{T}_1) - SSR_1(\hat{\tau}_1|\hat{T}_1)}{SSR_1(\hat{T}_1)/(\hat{T}_1 - p)}, \frac{SSR_2(\hat{T}_1) - SSR_2(\hat{\tau}_2|\hat{T}_1)}{SSR_2(\hat{T}_1)/(T - \hat{T}_1 - p)}\right), \tag{16}$$

$$SSR_1(\hat{T}_1) = \sum_{t=1}^{\hat{T}_1}\left(y_t - \hat{\boldsymbol{w}}_t'\hat{\boldsymbol{\beta}}_{(1)}\right)^2, \qquad SSR_2(\hat{T}_1) = \sum_{t=\hat{T}_1+1}^{T}\left(y_t - \hat{\boldsymbol{w}}_t'\hat{\boldsymbol{\beta}}_{(2)}\right)^2, \tag{17}$$

where $\hat{\boldsymbol{\beta}}_{(1)}$ is the 2SLS estimator calculated over the subsample $[1, \hat{T}_1]$ and $\hat{\boldsymbol{\beta}}_{(2)}$ is the 2SLS estimator calculated over the subsample $[\hat{T}_1 + 1, T]$.

To test $H_0 : m = 1$ versus $H_1 : m = 2$ via the Wald principle, the sup-$Wald$ test statistic is:

$$\sup\text{-}Wald_T(2|1) = \max\left(Wald_{T,1}(2|1;\hat{T}_1),\ Wald_{T,2}(2|1;\hat{T}_1)\right), \tag{18}$$

where $Wald_{T,1}(2|1;\hat{T}_1)$ and $Wald_{T,2}(2|1;\hat{T}_1)$ are the sup-$Wald$ statistics for testing for zero breaks against one break in the sub-sample before $\hat{T}_1$ and after $\hat{T}_1$ respectively:

$$Wald_{T,1}(2|1;\hat{T}_1) = \sup_{\tau_1\in\boldsymbol{\Lambda}_1} T\hat{\boldsymbol{\beta}}(\eta_1)'\boldsymbol{R}_1'[\boldsymbol{R}_1\hat{\boldsymbol{V}}(\eta_1)\boldsymbol{R}_1']^{-1}\boldsymbol{R}_1\hat{\boldsymbol{\beta}}(\eta_1), \tag{19}$$

$$Wald_{T,2}(2|1;\hat{T}_1) = \sup_{\tau_2\in\boldsymbol{\Lambda}_2} T\hat{\boldsymbol{\beta}}(\eta_2)'\boldsymbol{R}_1'[\boldsymbol{R}_1\hat{\boldsymbol{V}}(\eta_2)\boldsymbol{R}_1']^{-1}\boldsymbol{R}_1\hat{\boldsymbol{\beta}}(\eta_2), \tag{20}$$

where $\boldsymbol{R}_1$ is defined as $\boldsymbol{R}_k$ in case (a) but with $k = 1$ (the alternative hypothesis of an additional break in SE), and $\hat{\boldsymbol{\beta}}(\eta_i) = \left(\hat{\boldsymbol{\beta}}'_{(1)}(\eta_i), \hat{\boldsymbol{\beta}}'_{(2)}(\eta_i)\right)'$. For the candidate second break in SE located before the first break in SE we denote $\hat{\boldsymbol{\beta}}(\eta_1) = \left(\hat{\boldsymbol{\beta}}'_{(1)}(\eta_1), \hat{\boldsymbol{\beta}}'_{(2)}(\eta_1)\right)'$ where $\hat{\boldsymbol{\beta}}_{(1)}(\eta_1)$ and $\hat{\boldsymbol{\beta}}_{(2)}(\eta_1)$ are the 2SLS estimators in the sub-samples $I_1(\eta_1) = [1, \tau_1]$ and $I_2(\eta_1) = [\tau_1 + 1, \hat{T}_1]$ respectively. For the candidate second break in SE located after the

first break in SE we denote $\hat{\boldsymbol{\beta}}(\eta_2) = \left(\hat{\boldsymbol{\beta}}'_{(1)}(\eta_2), \hat{\boldsymbol{\beta}}'_{(2)}(\eta_2)\right)'$ where $\hat{\boldsymbol{\beta}}_{(1)}(\eta_2)$ and $\hat{\boldsymbol{\beta}}_{(2)}(\eta_2)$ are the 2SLS estimators in the sub-samples $I_1(\eta_2) = [\hat{T}_1 + 1, \hat{T}_1 + \tau_2]$ and $I_2(\eta_2) = [\tau_2 + 1, T]$ respectively. In addition, in (19), we let:

$$\hat{\boldsymbol{Q}}_{(i)}(\eta_j) = T^{-1} \sum_{t \in I_i(\eta_j)} \hat{\boldsymbol{w}}_t \hat{\boldsymbol{w}}'_t\,, \ \hat{\boldsymbol{V}}(\eta_i) = \text{diag}\left(\hat{\boldsymbol{V}}_{(1)}(\eta_i), \hat{\boldsymbol{V}}_{(2)}(\eta_i)\right)\,, \ \hat{\boldsymbol{V}}_{(i)}(\eta_j) \ = \ \hat{\boldsymbol{Q}}^{-1}_{(i)}(\eta_j)\ \hat{\boldsymbol{M}}_{(i)}(\eta_j)\ \hat{\boldsymbol{Q}}^{-1}_{(i)}(\eta_j), \quad (21)$$

$$\hat{\boldsymbol{M}}_{(i)}(\eta_j) \xrightarrow{p} \lim_{T\to\infty} \text{var}\left(T^{-1/2} \sum_{t \in I_i(\eta_j)} \boldsymbol{\Upsilon}^{0'}(t,T)\boldsymbol{z}_t \left(u_t + \boldsymbol{v}'_t \boldsymbol{\beta}^0_{x,(i)}\right)\right)\,, \ i,j = 1,2. \quad (22)$$

3 Assumptions

In this section we introduce the assumptions needed to derive the limiting distribution of the sup-F and sup-$Wald$ tests introduced in the previous section.

Assumption 1. *(i) Let* $\boldsymbol{\varepsilon}_t = (u_t, \boldsymbol{v}'_t)'$. *Then* $E(\boldsymbol{\varepsilon}_s \boldsymbol{r}'_t) = 0$ *for all* $t, s = 1, \dots T$, *and* $\text{E}(\boldsymbol{\varepsilon}_{t-l}\boldsymbol{r}'_t | \mathcal{F}^r_{t-l}) = 0$ *for all* t *and* $1 \le l \le t-1$, *and* $\mathcal{F}^r_t$ *is the* σ*-algebra generated by* $\{\boldsymbol{r}_t, \boldsymbol{r}_{t-1}, \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{r}_{t-2}, \boldsymbol{\varepsilon}_{t-2}, \dots\}$.

(ii) $\text{E}(\boldsymbol{r}_{t-l}\boldsymbol{r}'_{t-\kappa}) = \boldsymbol{\Gamma}_{|l-\kappa|}$ *for all* t *and* $l, \kappa > 0$.

Assumption 1(i) states that the errors are uncorrelated with past or future values of the exogenous regressors $\boldsymbol{r}_t$, and that we treat $\boldsymbol{r}_t$ as given in all bootstraps. It is not a strong assumption given that we already include lags of the dependent and the endogenous variables in the SE from (1).[12] Assumptions 1(ii) is a stationarity assumption useful in the bootstrap section.

We assume the following about the true break fractions and true parameters in each regime.

Assumption 2. $T^0_i = [T\lambda^0_i]$, *where* $0 < \lambda^0_1 < \ldots < \lambda^0_m < 1$.

Assumption 3. $\boldsymbol{\beta}^0_{(i+1)} - \boldsymbol{\beta}^0_{(i)} = \boldsymbol{\nu}^0_{i,T} \ = \ \boldsymbol{\nu}^0_i r^*_T$, *where* $r^*_T = T^{-\alpha}$, $0 < \alpha < 0.5$, $i = 1, \dots, m+1$ *and* $\boldsymbol{\nu}^0_i$ *is a vector of constants. Let* $\boldsymbol{\beta}^0 = \left(\boldsymbol{\beta}^{0\prime}_{\boldsymbol{x}}, \boldsymbol{\beta}^{0\prime}_{\boldsymbol{z}_1}\right)'$ *be the common limiting value of* $\boldsymbol{\beta}^0_{(i)} = \left(\boldsymbol{\beta}^{0\prime}_{\boldsymbol{x},(i)}, \boldsymbol{\beta}^{0\prime}_{\boldsymbol{z}_1,(i)}\right)'$.

Assumption 4. $T^*_j = [T\pi^0_j]$, *where* $0 < \pi^0_1 < \ldots < \pi^0_h < 1$.

Assumption 5. $\boldsymbol{\Delta}^0_{(j+1)} - \boldsymbol{\Delta}^0_{(j)} = \boldsymbol{S}^*_{j,T} \ = \ \boldsymbol{S}^0_j s^*_T$ *where* $s^*_T = T^{-\rho}$, $0 < \rho < 0.5$, $j = 1, \dots, h+1$ *and* $\boldsymbol{S}^0_j$ *a matrix of constants. Let* $\boldsymbol{\Delta}^0$ *be the common limiting value of* $\boldsymbol{\Delta}^0_{(j)}$, *where* $\text{rk}(\boldsymbol{\Delta}^0) = p_1$.

Assumption 6. *The minimization in (8) is over all partitions* $\boldsymbol{\lambda}$ *such that* $T_i - T_{i-1} = [T\lambda_i] - [T\lambda_{i-1}] > \max(q-1, \epsilon T)$ *for some* $\epsilon > 0$ *and* $\epsilon < \min_i(\lambda^0_{i+1} - \lambda^0_i)$ *and* $\epsilon < \min_i(\pi^0_{i+1} - \pi^0_i)$.

Assumption 2 implies the break points are asymptotically distinct in the SE, and Assumption 4 implies the break points are asymptotically distinct in the RF. However, there can be common break points in the SE and RF. By Assumption 3, $\boldsymbol{\beta}^0_{(i)} = \boldsymbol{\beta}^0 + O(T^{-\alpha})$ and by Assumption 5, $\boldsymbol{\Delta}^0_{(j)} = \boldsymbol{\Delta}^0 + O(T^{-\rho})$, where $\boldsymbol{\beta}^0, \boldsymbol{\Delta}^0$ are the common limiting values of $\boldsymbol{\beta}^0_{(i)}$ and $\boldsymbol{\Delta}^0_{(j)}$, so we assume shrinking breaks, but we allow for the RF breaks to be smaller or larger than the ones in the SE.[13] Assumption 6 requires that each segment considered in the minimization contains a positive fraction of the sample asymptotically; in practice ϵ is chosen to be small so the last part of the assumption holds. All these assumptions are standard in the break point literature.

Since SE (1) and RF (3) form a dynamic model, we also need stability assumptions. To introduce these stability assumptions, we consider **Case (I)** when there are no SE breaks ($m = 0$), but there are RF breaks

[12] If, for example, $\text{E}(\boldsymbol{\varepsilon}_t \boldsymbol{r}_{t-j}) \neq 0$, for $j > 0$, then any bootstrap method would have to mimic this correlation.

[13] We can allow for fixed (large) breaks instead of shrinking (moderate) breaks. Our simulations show that the bootstrap works well in both cases, but our theory is developed under shrinking breaks because the notation and the proofs are greatly simplified.

($h > 0$), and **Case (II)** when there are breaks in both SE and RF. For clarity, we focus on the case when there is one SE break ($m = 1$) and there is at least one RF break ($h > 0$); the SE break can coincide with one of the RF breaks. However, a closer inspection of our proofs shows that our results also extend to the case when $m > 1$. To present the stability conditions, we derive an alternative representation for the data generation process under **Cases (I)** and **(II)**. It is shown in Appendix A below that we can write (1) and (3) in the VARX($\tilde{p}$,0) form:[14]

$$\underbrace{\tilde{\boldsymbol{y}}_t}_{(p_1+1)\times 1} = \sum_{i=1}^{\tilde{p}} \underbrace{\boldsymbol{C}_i(t,T)}_{(p_1+1)\times(p_1+1)} \times \underbrace{\tilde{\boldsymbol{y}}_{t-i}}_{(p_1+1)\times 1} + \underbrace{\boldsymbol{J}(t,T)}_{(p_1+1)\times(q_1+q_2)} \times \underbrace{\boldsymbol{r}_t}_{(q_1+q_2)\times 1} + \underbrace{\boldsymbol{e}_t}_{(p_1+1)\times 1}, \tag{23}$$

where $\tilde{\boldsymbol{y}}_{t-i} = (y_{t-i}, \boldsymbol{x}'_{t-i})'$ was defined in (5), $\boldsymbol{C}_i(t,T)$ and $\boldsymbol{J}(t,T)$ are matrices of coefficients given by (64) for **Case (I)** and by (68) for **Case (II)** in Appendix A below. For **Case (I)**, $\boldsymbol{e}_t = \boldsymbol{A}_0^{-1}\boldsymbol{\varepsilon}_t$, while for **Case (II)**, $\boldsymbol{e}_t = \left(\boldsymbol{A}_{0,(1)}^{-1} 1_{t\in[1,T_1^0]} + \boldsymbol{A}_{0,(2)}^{-1} 1_{t\in[T_1^0+1,T]}\right)\boldsymbol{\varepsilon}_t$, where $\boldsymbol{\varepsilon}_t = (u_t, \boldsymbol{v}'_t)'$ was defined in Assumption 1, and $\boldsymbol{A}_0$, respectively $\boldsymbol{A}_{0,(1)}$, $\boldsymbol{A}_{0,(2)}$ are $(p_1+1)\times(p_1+1)$ matrices of coefficients for **Case (I)**, respectively **Case (II)**; see (63) and (66) in Appendix A. In Appendix A, we denote by $\boldsymbol{C}_i$ the common limiting value of $\boldsymbol{C}_i(t,T)$ for **Cases (I)** and **(II)**. We are now in position to state the following stability assumptions.

Assumption 7. $\left|\boldsymbol{I}_{p_1+1} - \boldsymbol{C}_1(t,T)a - \boldsymbol{C}_2(t,T)a^2 - \cdots - \boldsymbol{C}_{\tilde{p}}(t,T)a^{\tilde{p}}\right| \neq \boldsymbol{0}_{p_1+1}$, *for all* $t = 1,\ldots,T$, *and for all* $|a| \leq 1$.

Assumption 8. $\left|\boldsymbol{I}_{p_1+1} - \boldsymbol{C}_1 a - \boldsymbol{C}_2 a^2 - \cdots - \boldsymbol{C}_{\tilde{p}} a^{\tilde{p}}\right| \neq \boldsymbol{0}_{p_1+1}$, *for all* $|a| \leq 1$.

We allow the errors in (1) and (3) to be serially correlated and heteroskedastic via the following assumption.

Assumption 9. *(i)* $\boldsymbol{h}_t = \boldsymbol{\varepsilon}_t \otimes \boldsymbol{z}_t$ *is an array of real valued* $\tilde{q}\times 1$ *random vectors (where* $\tilde{q} = (p_1+1)q$*) defined on the probability space* $(\Omega, \mathcal{F}, P)$, *and* $\boldsymbol{V}_T = \text{var}\left(T^{-1/2}\sum_{t=1}^T \boldsymbol{h}_t\right)$ *has eigenvalues* $\gamma_{T,j} = O(1)$, $j = 1,\ldots,\tilde{q}$;

(ii) $\text{E}(h_{t,i}) = 0$ *and* $\sup_t \|h_{t,i}\|_d < \kappa < \infty$ *for some* $d > 4$, *for all* t *and for* $i = 1,\ldots,\tilde{q}$, *where* $h_{t,i}$ *is the* i^{th} *element of* $\boldsymbol{h}_t$;

(iii) $\{h_{t,i}\}$ *is near epoch dependent with respect to the mixing process* $\{g_t\}$ *such that* $\|h_{t,i} - \text{E}(h_{t,i}|\mathcal{G}_{t-n}^{t+n})\|_2 \leq \nu_n$ *with* $\nu_n = O(n^{-1/2})$, *where* $\mathcal{G}_{t-n}^{t+n}$ *is a* σ*-algebra based on* $(g_{t-n},\ldots,g_{t+n})$;

(iv) $\{g_t\}$ *is either* ϕ*-mixing of size* $n^{-d/(2(d-1))}$ *or* α*-mixing of size* $n^{-d/(d-2)}$.

Assumption 9 allows for substantial dependence and heterogeneity in $\boldsymbol{h}_t$ but also imposes sufficient restrictions to allow for a functional central limit theorem for $T^{-1/2}\sum_{t=1}^{[Ts]} \boldsymbol{h}_t$; see Wooldridge and White (1988), Theorem 2.11.

We also assume that the instruments $\boldsymbol{z}_t$ in (4) are strong via a standard rank condition for identification in IV estimation of linear models:

Assumption 10. $\text{rk}\left(\boldsymbol{\Upsilon}_{(i)}^0\right) = p$, *where* $\boldsymbol{\Upsilon}_{(i)}^0 = \left(\boldsymbol{\Delta}_{(i)}^0, \boldsymbol{\Pi}\right)$ *is a matrix of size* $q \times p$, *for* $i = 1,2,\ldots,h+1$, $\boldsymbol{\Pi}' = (\boldsymbol{I}_{p_2}, \boldsymbol{O}_{p_2,q-p_2})$. *Let* $\boldsymbol{\Upsilon}^0$ *be the common limiting value of the* $\boldsymbol{\Upsilon}_{(i)}^0$*'s.*[15]

By Assumption 5, $\boldsymbol{\Upsilon}_{(i)}^0 = \boldsymbol{\Upsilon}^0 + O(T^{-\rho})$. By Assumptions 4-8, 10 and Theorem 2 in HHB, it follows that $\hat{\pi}_j - \pi_j^0 = O_p(T^{2\rho-1})$, $0 < \rho < 0.5$, $j = 1,\ldots,h$. In addition, we show in the Supplemental Appendix (Lemma 1.2) that $\hat{\lambda}_1 - \lambda_1^0 = O_p(T^{2\alpha-1})$, $0 < \alpha < 0.5$, $i = 1,\ldots,m+1$. It follows that the SE and RF break fractions are

[14] We set the number of lags for $\boldsymbol{r}_t$ to zero, but $\boldsymbol{r}_t$ could involve lags at the expense of additional notation.

[15] The notation $\boldsymbol{\Upsilon}_{(i)}^0$ is convenient for calculations involving the augmented vector of projected endogenous regressors and observed exogenous regressors in the second stage estimation.

converging fast enough so that the asymptotic distribution of the parameter estimates (which we use to derive the asymptotic distributions of the multiple break tests below) is unaffected by the randomness in the break fraction estimates.

To develop tests for SE breaks, we need to impose further restrictions on the instrument cross-product matrix and on the long-run variance of $\boldsymbol{h}_t$. Similar to Assumption A8 of BP, we impose Assumption 11 which allows for shrinking breaks in the marginal distribution of the regressors and excludes the presence of trending regressors.

Assumption 11. $T^{-1}\sum_{t=1}^{[Ts]} \boldsymbol{z}_{t-l}\boldsymbol{z}'_{t-l} \xrightarrow{p} s\boldsymbol{Q}_{\boldsymbol{zz}}$ *uniformly in* $s \in [0,1]$ *for all* $l \geq 0$*, where* $\boldsymbol{Q}_{\boldsymbol{zz}}$ *is a* $q \times q$ *positive definite (hereafter pd) matrix of constants.*

For the sup-F test, as usual, we need that the errors are serially uncorrelated and conditionally homoskedastic:

Assumption 12. *Let* $\mathcal{F}_t$ *be the* σ*-field generated by* $\boldsymbol{z}_t, \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{z}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \boldsymbol{z}_{t-2}, \ldots$*. Then:*

(i) $\mathrm{E}(\boldsymbol{\varepsilon}_{t-l} \mid \mathcal{F}_{t-l}) = 0$*, almost surely, for all* $t, l > 0$*.*

(ii) $\mathrm{E}(\boldsymbol{\varepsilon}_{t-l}\boldsymbol{\varepsilon}'_{t-l}) = \mathrm{E}(\boldsymbol{\varepsilon}_{t-l}\boldsymbol{\varepsilon}'_{t-l} \mid \mathcal{F}_{t-l}) = \boldsymbol{\Omega}$ *for all* $t, l > 0$*, with* $\boldsymbol{\Omega} = \begin{pmatrix} \sigma_u^2 & \boldsymbol{\sigma}'_{u\boldsymbol{v}} \\ \boldsymbol{\sigma}_{u\boldsymbol{v}} & \boldsymbol{\Omega}_{\boldsymbol{v}} \end{pmatrix}$,

where $\boldsymbol{\Omega}$ *and* $\boldsymbol{\Omega}_{\boldsymbol{v}}$ *are* $(p_1+1)\times(p_1+1)$*, respectively* $p_1 \times p_1$ *pd matrices of constants, and* $\boldsymbol{\sigma}_{u\boldsymbol{v}}$ *is* $p_1 \times 1$ *vector of constants.*

For statistics based on the Wald principle, we allow for serial correlation and heteroskedasticity. We first state the assumptions and then discuss them.

Assumption 13.

(i) $\lim_{T\to\infty} \boldsymbol{V}_T(s) = \lim_{T\to\infty} T^{-1}\operatorname{var}\left(\sum_{t=1}^{[Ts]} \boldsymbol{h}_t\right) = s\boldsymbol{\Sigma}$ *for all* t *uniformly in* $s \in [0,1]$*, with*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_u & \boldsymbol{\Sigma}_{u\boldsymbol{v}} \\ \boldsymbol{\Sigma}'_{u\boldsymbol{v}} & \boldsymbol{\Sigma}_{\boldsymbol{v}} \end{pmatrix},$$

where $\boldsymbol{\Sigma}$*,* $\boldsymbol{\Sigma}_u$ *and* $\boldsymbol{\Sigma}_{\boldsymbol{v}}$ *are* $\tilde{q}\times\tilde{q}$*,* $q\times q$*, respectively* $(p_1 q)\times(p_1 q)$ *pd matrices of constants.*

(ii) $\mathrm{E}(\boldsymbol{h}_t\boldsymbol{h}_l) = 0$ *for* $t \neq l$*, and* $\lim_{T\to\infty} \boldsymbol{V}_T(s) = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{[Ts]} \mathrm{E}(\boldsymbol{h}_t\boldsymbol{h}'_t) = s\tilde{\boldsymbol{\Sigma}}$ *uniformly in* $s \in [0,1]$*, with*

$$\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \tilde{\boldsymbol{\Sigma}}_u & \tilde{\boldsymbol{\Sigma}}_{u\boldsymbol{v}} \\ \tilde{\boldsymbol{\Sigma}}'_{u\boldsymbol{v}} & \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}} \end{pmatrix},$$

where $\tilde{\boldsymbol{\Sigma}}_u$*,* $\tilde{\boldsymbol{\Sigma}}_{u\boldsymbol{v}}$ *and* $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}}$ *are* $\tilde{q}\times\tilde{q}$*,* $q\times q$*, respectively* $(p_1 q)\times(p_1 q)$ *pd matrices of constants.*

Assumption 13(i) is sufficient for deriving the asymptotic distribution of the sup-$Wald$ test. It refers to $\boldsymbol{h}_t = \boldsymbol{z}_t \otimes \boldsymbol{\varepsilon}_t$ and allows for serial correlation and unconditional hetoroskedasticity in $\boldsymbol{h}_t$, and as a consequence in the disturbances $\boldsymbol{\varepsilon}_t$ (because we assume that $\boldsymbol{z}_t$ contains an intercept). If serial correlation is present, then no lagged dependent variables are allowed in $\boldsymbol{z}_t$. Assumption 13(ii) states that $\boldsymbol{\Sigma}$ in Assumption 13(i) changes to $\tilde{\boldsymbol{\Sigma}}$ in the absence of serial correlation in $\boldsymbol{h}_t$.

4 Limiting distributions of sup-F and sup-$Wald$ tests

When there are no RF breaks ($h = 0$) and the RF is estimated over the full sample, HHB show that the test statistics introduced in Section 2 converge to the usual distributions tabulated in BP, and our theorems below

include this case. However, if there are breaks in RF ($h \neq 0$), then these test statistics no longer converge to the BP distributions. This has been noted in HHB, but we provide a formal proof here. The asymptotic distributions we derive in this section are useful for two other purposes: first, to prove the validity of the bootstrap - in Section 5.2, and second, to compare the performance of the tests based on asymptotic critical values with that of the bootstrap tests - in Section 6.

The limiting distribution of sup-$F_T(\boldsymbol{\lambda})$ for $k = 1, 2$, and $h = 1, 2, \ldots, h_{\max}$, under $H_0 : \boldsymbol{\beta}_{(i)}^0 = \boldsymbol{\beta}^0$, will be given by a Gaussian process that we need to define. To that end, note that under Assumptions 9, 11, and 12, by the multivariate functional central limit theorem in Wooldridge and White (1988), Theorem 2.11, we have:

$$T^{-1/2} \sum_{t=1}^{[Ts]} \boldsymbol{h}_t \Rightarrow (\boldsymbol{\Omega}^{1/2} \otimes \boldsymbol{Q}_{\boldsymbol{zz}}^{1/2}) \boldsymbol{W}(s), \tag{24}$$

where $\boldsymbol{\Omega}^{1/2}$ and $\boldsymbol{Q}_{\boldsymbol{zz}}^{1/2}$ are the matrix square roots of $\boldsymbol{\Omega}$, respectively $\boldsymbol{Q}_{\boldsymbol{zz}}$, and $\boldsymbol{W}(s) = \left(\boldsymbol{W}_1'(s), \ldots, \boldsymbol{W}_{p_1+1}'(s)\right)'$, where $\boldsymbol{W}(s)$ and $\boldsymbol{W}_i(s)$ are $\tilde{q} \times 1$ and $q \times 1$ vectors of standard independent Brownian motions. Let $\boldsymbol{n}_1$ and $\boldsymbol{N}_2$ be a $(p_1 + 1) \times 1$ vector and a $(p_1 + 1) \times p_1$ matrix respectively, such that $\boldsymbol{n}_1'\boldsymbol{n}_1 = \sigma_u$, $\boldsymbol{n}_1'\boldsymbol{N}_2 = \boldsymbol{\sigma}_{u\boldsymbol{v}}'$ and $\boldsymbol{N}_2'\boldsymbol{N}_2 = \boldsymbol{\Omega}_{\boldsymbol{v}}$. Then, by arguments similar to HHB, Supplemental Appendix, p.22, we have:

$$T^{-1/2} \sum_{t=1}^{[Ts]} \boldsymbol{z}_t u_t \Rightarrow (\boldsymbol{n}_1' \otimes \boldsymbol{Q}_{\boldsymbol{zz}}^{1/2}) \boldsymbol{W}(s), \qquad T^{-1/2} \sum_{t=1}^{[Ts]} \boldsymbol{z}_t \boldsymbol{v}_t' \Rightarrow \boldsymbol{Q}_{\boldsymbol{zz}}^{1/2} \tilde{\boldsymbol{W}}(s) \boldsymbol{N}_2, \tag{25}$$

with $\tilde{\boldsymbol{W}}(s) = (\boldsymbol{W}_1(s), \ldots, \boldsymbol{W}_{p_1+1}(s))$, a $q \times (p_1 + 1)$ matrix. Letting $\boldsymbol{D}(\cdot) \equiv (\boldsymbol{n}_1' \otimes \boldsymbol{I}_q) \boldsymbol{W}(\cdot)$ and $\boldsymbol{D}^*(\cdot) \equiv \tilde{\boldsymbol{W}}(\cdot) \boldsymbol{N}_2 \boldsymbol{\beta}_{\boldsymbol{x}}^0$, both $q \times 1$ vectors, we define the following processes below.

Definition 1. *For any random process $\boldsymbol{A}(\cdot) : [0, 1] \Rightarrow \mathbb{R}^d$, and a k-partition $\boldsymbol{\lambda}$, let*

$$\boldsymbol{K}_{\boldsymbol{A}}(\boldsymbol{\lambda}) = \left(\boldsymbol{A}'(\lambda_1), \boldsymbol{A}'(\lambda_2) - \boldsymbol{A}'(\lambda_1), \ldots, \boldsymbol{A}'(\lambda_i) - \boldsymbol{A}'(\lambda_{i-1}), \ldots, \boldsymbol{A}'(1) - \boldsymbol{A}'(\lambda_k)\right)'. \tag{26}$$

Definition 2. *Let $\boldsymbol{K}^*(\boldsymbol{\lambda}; \boldsymbol{\pi}^0) = \boldsymbol{K}_{\boldsymbol{D}}(\boldsymbol{\lambda}) + \boldsymbol{K}_{\boldsymbol{D}^*}(\boldsymbol{\lambda}) - (\boldsymbol{P}(\boldsymbol{\lambda}, \boldsymbol{\pi}^0) \otimes \boldsymbol{I}_q)\, \boldsymbol{K}_{\boldsymbol{D}^*}(\boldsymbol{\pi}^0)$, a $q(k+1) \times 1$ random vector, with $\boldsymbol{P}(\boldsymbol{\lambda}, \boldsymbol{\pi}^0)$ described in Definition B1 of the Appendix B, and $\boldsymbol{I}_q$ the $q \times q$ identity matrix. Also, let $\boldsymbol{K}^{**}(\boldsymbol{\lambda}; \boldsymbol{\pi}^0)$ be computed from $\boldsymbol{K}^*(\boldsymbol{\lambda}; \boldsymbol{\pi}^0)$ by selecting and stacking, in order, from each block $q \times 1$, the first p elements.*

With these definitions, the asymptotic distribution of the sup-F_T test is stated below.

Theorem 1. *Under Assumptions 1 and 4-12, the test* sup-F_T (10) *for testing for $m = 0$ breaks in SE against $m = k$ breaks in SE is:*

$$\sup\text{-}F_T \Rightarrow \sup_{\lambda \in \boldsymbol{\Lambda}_\epsilon} \left\{ \frac{1}{p\ \sigma_1^2} \boldsymbol{K}^{**\prime}(\boldsymbol{\lambda}; \boldsymbol{\pi}^0) \left(\boldsymbol{H}^*(\boldsymbol{\lambda}) \otimes \boldsymbol{I}_p\right) \boldsymbol{K}^{**}(\boldsymbol{\lambda}; \boldsymbol{\pi}^0) \right\}, \tag{27}$$

where $\sigma_1^2 = \sigma_u^2 + 2\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime} \boldsymbol{\sigma}_{u\boldsymbol{v}} + \boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime} \boldsymbol{\Omega}_{\boldsymbol{v}} \boldsymbol{\beta}_{\boldsymbol{x}}^0$, and $\boldsymbol{H}^*(\boldsymbol{\lambda})$ is the $k \times k$ matrix with diagonal elements equal to $(1 - \lambda_i + \lambda_{i-1})/(\lambda_i - \lambda_{i-1})$, $i = 1, \ldots, k$, and the rest of the elements equal to -1.

Theorem 1 characterizes the limiting distribution of sup-F_T for any number of breaks $h = 0, 1, 2, \ldots, h_{\max}$ in RF, but the specific form of the distribution depends on h and $\boldsymbol{\pi}$. To see the intuition behind (27), note that in the special case with only exogenous regressors, we have that $\boldsymbol{x}_t = \boldsymbol{v}_t = \boldsymbol{0}_{p_1}$, so we set $\boldsymbol{z}_{1,t} = \boldsymbol{z}_t$ and $\boldsymbol{\sigma}_{u\boldsymbol{v}} = \boldsymbol{0}_{p_1}$. Therefore, $\boldsymbol{K}^*(\boldsymbol{\lambda}; \boldsymbol{\pi}^0) = \boldsymbol{K}_{\boldsymbol{D}}(\boldsymbol{\lambda})$, and the distribution of the sup-F_T in Theorem 1 is the same as in BP, since it is generated exclusively by partial sums of $\{\boldsymbol{z}_{1,t}\ u_t\}$. In the other special case where some regressors are endogenous but the RF is stable, $h = 0$ and we set $\boldsymbol{\pi}^0 = 1$. Hence, $\boldsymbol{P}(\boldsymbol{\lambda}, \boldsymbol{\pi}^0) = (\lambda_1, \ldots, \lambda_k)'$, and $\boldsymbol{K}_{\boldsymbol{D}^*}(\boldsymbol{\pi}^0) = \boldsymbol{K}_{\boldsymbol{D}^*}(1)$. Therefore, $\boldsymbol{K}^*(\boldsymbol{\lambda}; \boldsymbol{\pi}^0) = \boldsymbol{K}_{\boldsymbol{D}}(\boldsymbol{\lambda}) + \boldsymbol{K}_{\boldsymbol{D}^*}(\boldsymbol{\lambda}) - ((\lambda_1, \ldots, \lambda_k)' \otimes \boldsymbol{I}_q)\ \boldsymbol{K}_{\boldsymbol{D}^*}(1)$, and it can be

shown that the distribution of the sup-F_T in Theorem 1 is the same as in BP, but it is generated instead by the partial sums of $\boldsymbol{z}_t(u_t + \boldsymbol{v}_t\boldsymbol{\beta}_x^0)$ (this is also the result in HHB). However, when there are RF breaks ($h \neq 0$), then this distribution is not pivotal, since it depends on the number of RF breaks and their location, as indicated by the term $(\boldsymbol{P}(\boldsymbol{\lambda}, \boldsymbol{\pi}^0) \otimes \boldsymbol{I}_q)\ \boldsymbol{K}_{\boldsymbol{D}^*}(\boldsymbol{\pi}^0)$. HHB side-step this issue by splitting the RF into stable subsamples, and proceeding with the tests for stable RF in these subsamples. However, given that these subsamples can be small, further break point tests in sub-samples may be highly inaccurate. Therefore, in this paper, we propose bootstrapping this test and other structural change tests over the full sample.

In the presence of autocorrelation and heteroskedasticity, we employ the sup-$Wald_T$ test instead of the sup-F_T. Similar to sup-F_T, the asymptotic null distribution of the sup-$Wald_T$ test is a Gaussian process that we need to define. To that end, note that under Assumptions 9, 11, and 13(i), by the multivariate functional central limit theorem in Wooldridge and White (1988), Theorem 2.11, we have:

$$T^{-1/2}\sum_{t=1}^{[Ts]} \boldsymbol{h}_t \Rightarrow \boldsymbol{\Sigma}^{1/2}\boldsymbol{W}(s), \tag{28}$$

where $\boldsymbol{\Sigma}$ is defined in Assumption 13(i). Let $\tilde{\boldsymbol{N}}_1$ and $\tilde{\boldsymbol{N}}_2$ be $\tilde{q} \times q$, respectively $\tilde{q} \times p_1 q$ matrices such that $\tilde{\boldsymbol{N}}_1'\tilde{\boldsymbol{N}}_1 = \boldsymbol{\Sigma}_u$, $\tilde{\boldsymbol{N}}_1'\tilde{\boldsymbol{N}}_2 = \boldsymbol{\Sigma}_{u\boldsymbol{v}}$ and $\tilde{\boldsymbol{N}}_2'\tilde{\boldsymbol{N}}_2 = \boldsymbol{\Sigma}_{\boldsymbol{v}}$. Let $\tilde{\boldsymbol{N}}_2' = \left(\tilde{\boldsymbol{N}}_{2,1}', \ldots, \tilde{\boldsymbol{N}}_{2,p_1}'\right)$, where $\tilde{\boldsymbol{N}}_{2,i}'$ is of size $q \times \tilde{q}$, for $i = 1, \ldots, p_1$. Then:

$$T^{-1/2}\sum_{t=1}^{[Ts]} \boldsymbol{z}_t u_t \Rightarrow \tilde{\boldsymbol{N}}_1'\boldsymbol{W}(s), \qquad T^{-1/2}\sum_{t=1}^{[Ts]} \boldsymbol{z}_t \boldsymbol{v}_t' \Rightarrow \tilde{\tilde{\boldsymbol{W}}}(s), \tag{29}$$

where $\tilde{\tilde{\boldsymbol{W}}}(s) = \left(\tilde{\boldsymbol{N}}_{2,1}'\boldsymbol{W}(s), \ldots, \tilde{\boldsymbol{N}}_{2,p_1}'\boldsymbol{W}(s)\right)$ is a $q \times p_1$ matrix of Brownian motions. Letting the $q \times 1$ vector processes $\tilde{\boldsymbol{D}}(\cdot) \equiv \tilde{\boldsymbol{N}}_1'\boldsymbol{W}(\cdot)$, $\tilde{\boldsymbol{D}}^*(\cdot) \equiv \tilde{\tilde{\boldsymbol{W}}}(\cdot)\boldsymbol{\beta}_{\boldsymbol{x}}^0$, we define the following process below.

Definition 3. *Let* $\tilde{\boldsymbol{K}}^*(\boldsymbol{\lambda}; \boldsymbol{\pi}^0) = \boldsymbol{K}_{\tilde{\boldsymbol{D}}}(\boldsymbol{\lambda}) + \boldsymbol{K}_{\tilde{\boldsymbol{D}}^*}(\boldsymbol{\lambda}) - (\boldsymbol{P}(\boldsymbol{\lambda}, \boldsymbol{\pi}^0) \otimes \boldsymbol{I}_q)\ \boldsymbol{K}_{\tilde{\boldsymbol{D}}^*}(\boldsymbol{\pi}^0)$, *with* $\boldsymbol{P}(\boldsymbol{\lambda}, \boldsymbol{\pi}^0)$ *in Definition B1.*

Then the limiting distribution of the sup-$Wald_T$ test is stated below.

Theorem 2. *Under Assumptions 1, 4-11, and 13(i), the asymptotic distribution of* sup-$Wald_T$ *for testing for* $m = 0$ *breaks in SE against* $m = k$ *breaks in SE is:*

$$\text{sup-}Wald_T \Rightarrow \sup_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}_\epsilon} \left\{ \tilde{\boldsymbol{K}}^{*\prime}(\boldsymbol{\lambda}; \boldsymbol{\pi}^0)\ (\boldsymbol{H}^*(\boldsymbol{\lambda}) \otimes \boldsymbol{V}^*)\ \tilde{\boldsymbol{K}}^*(\boldsymbol{\lambda}; \boldsymbol{\pi}^0) \right\}, \tag{30}$$

with $\boldsymbol{V}^* = \boldsymbol{\Upsilon}^0(\boldsymbol{\Upsilon}^{0'}\boldsymbol{\Sigma}_1\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Upsilon}^{0'}$, *where* $\boldsymbol{\Upsilon}^0$ *is the common limit value of the* $\boldsymbol{\Upsilon}_{(i)}^0$*'s,* $i = 1, \ldots, h$, $\boldsymbol{\Sigma}_1$ *is as in Definition B2 of the Appendix B, and* $\boldsymbol{H}^*(\boldsymbol{\lambda})$ *is as in Theorem 1.*

Theorem 2 characterizes the limiting distribution of sup-$Wald_T$ for any number of breaks $h = 0, 1, 2, \ldots, h_{\max}$ in RF, but the specific form of the distribution depends on h and $\boldsymbol{\pi}$. Note that the distribution in Theorem 2 is the BP distribution when the regressors are exogenous or the RF is stable, but it depends on the number and location of the RF breaks when the RF is unstable.

Next, we derive the distribution of $F_T(2|1)$, based on the following Gaussian processes.

Definition 4. *Let* $\boldsymbol{K}_i^*(\eta_i; \boldsymbol{\pi}^0, \lambda_1^0) = \boldsymbol{K}_{\boldsymbol{D}}(\boldsymbol{s}_i) + \boldsymbol{K}_{\boldsymbol{D}^*}(\boldsymbol{s}_i) - (\boldsymbol{P}(\eta_i, \boldsymbol{\pi}^0) \otimes \boldsymbol{I}_q)\boldsymbol{K}_{\boldsymbol{D}^*}(\boldsymbol{\pi}^0)$, *where* $\boldsymbol{s}_1 = (0, \eta_1, \lambda_1^0)$ *is a partition of the interval* $[0, \lambda_1^0]$, $\boldsymbol{s}_2 = (\lambda_1^0, \lambda_1^0 + \eta_2, 1)$ *is a partition of the interval* $[\lambda_1^0 + 1, 1]$, *and* $\boldsymbol{P}(\eta_i, \boldsymbol{\pi}^0)$ *are defined in Definition B3 of the Appendix B, for* $i = 1, 2$. *Also, let* $\boldsymbol{K}_i^{**}(\eta_i; \boldsymbol{\pi}^0, \lambda_1^0)$ *be computed from* $\boldsymbol{K}_i^*(\eta_i; \boldsymbol{\pi}^0, \lambda_1^0)$ *by selecting and stacking, in order, from each* $q \times 1$ *block, the first* $p \times 1$ *elements.*

The asymptotic distribution of the test for one SE break against 2 SE breaks, sup-$F_T(2|1)$, is stated below.

Theorem 3. *Under Assumptions 1-12 the asymptotic distribution of* sup-F *under the null hypothesis* $m = 1$ *breaks in SE against the alternative* $m = 2$ *breaks in SE is:*

$$\sup\text{-}F_T(2|1) \Rightarrow \max\left(F_1(\eta_1;\boldsymbol{\pi}^0,\lambda_1^0), F_2(\eta_2;\boldsymbol{\pi}^0,\lambda_1^0)\right), \tag{31}$$

where

$$F_1(\eta_1;\boldsymbol{\pi}^0,\lambda_1^0) = \frac{1}{\sigma_1^2\lambda_1^0}\sup_{\tau_1\in\boldsymbol{\Lambda}_1}\left\{\boldsymbol{K}_1^{**'}(\eta_1;\boldsymbol{\pi}^0,\lambda_1^0)\left(\boldsymbol{H}_1^*(\eta_1)\otimes\boldsymbol{I}_p\right)\boldsymbol{K}_1^{**}(\eta_1;\boldsymbol{\pi}^0,\lambda_1^0)\right\},$$
$$F_2(\eta_2;\boldsymbol{\pi}^0,\lambda_1^0) = \frac{1}{\sigma_1^2\ (1-\lambda_1^0)}\sup_{\tau_2\in\boldsymbol{\Lambda}_2}\left\{\boldsymbol{K}_2^{**'}(\eta_2;\boldsymbol{\pi}^0,\lambda_1^0)\left(\boldsymbol{H}_2^*(\eta_2)\otimes\boldsymbol{I}_p\right)\boldsymbol{K}_2^{**}(\eta_2;\boldsymbol{\pi}^0,\lambda_1^0)\right\},$$

$$\boldsymbol{H}_1^*(\eta_1) = \begin{pmatrix} \frac{1-\eta_1}{\eta_1} & -1 \\ -1 & \frac{\eta_1}{1-\eta_1}\end{pmatrix},\ \boldsymbol{H}_2^*(\eta_2) = \begin{pmatrix} \frac{1-\eta_2}{\eta_2} & -1 \\ -1 & \frac{\eta_2}{1-\eta_2}\end{pmatrix}.$$

Theorem 3 characterizes the limiting distribution of sup-$F_T(2|1)$ for any number of breaks $h = 0, 1, 2, \ldots, h_{\max}$ in RF, but the specific form of the distribution depends on h and $\boldsymbol{\pi}$.

Next, we derive the distribution of $Wald_T(2|1)$, based on the following Gaussian processes.

Definition 5. *Let* $\tilde{\boldsymbol{K}}_i^*(\eta_i;\boldsymbol{\pi}^0,\lambda_1^0) = \boldsymbol{K}_{\tilde{\boldsymbol{D}}}(\boldsymbol{s}_i) + \boldsymbol{K}_{\tilde{\boldsymbol{D}}^*}(\boldsymbol{s}_i) - (\boldsymbol{P}(\eta_i,\boldsymbol{\pi}^0)\otimes\boldsymbol{I}_q)\boldsymbol{K}_{\tilde{\boldsymbol{D}}^*}(\boldsymbol{\pi}^0)$*, where* $\boldsymbol{s}_1 = (0,\eta_1,\lambda_1^0)$ *is a partition of the interval* $[0,\lambda_1^0]$*,* $\boldsymbol{s}_2 = (\lambda_1^0,\lambda_1^0+\eta_2,1)$ *is a partition of the interval* $[\lambda_1^0+1,1]$*, and* $\boldsymbol{P}(\eta_i,\boldsymbol{\pi}^0)$ *are defined in Definition B3, for* $i = 1, 2$*.*

With this definition, the asymptotic distribution of the sup-$Wald$ test for one SE break against two SE breaks, $Wald_T(2|1)$, is stated below.

Theorem 4. *Under Assumptions 1-11, 13(i) and* $H_0: m = 1$ *versus* $H_1: m = 2$*,*

$$\sup\text{-}Wald_T(2|1) \Rightarrow \max\left(Wald_1(\eta_1;\boldsymbol{\pi}^0,\lambda_1^0), Wald_2(\eta_2;\boldsymbol{\pi}^0,\lambda_1^0)\right), \tag{32}$$

where

$$Wald_1(\eta_1;\boldsymbol{\pi}^0,\lambda_1^0) = \frac{1}{\lambda_1^0}\sup_{\tau_1\in\boldsymbol{\Lambda}_1}\left\{\tilde{\boldsymbol{K}}_1^{*'}(\eta_1;\boldsymbol{\pi}^0,\lambda_1^0)\left(\boldsymbol{H}_1^*(\eta_1)\otimes\boldsymbol{V}^*\right)\ \tilde{\boldsymbol{K}}_1^*(\eta_1;\boldsymbol{\pi}^0,\lambda_1^0)\right\},$$
$$Wald_2(\eta_2;\boldsymbol{\pi}^0,\lambda_1^0) = \frac{1}{1-\lambda_1^0}\sup_{\tau_2\in\boldsymbol{\Lambda}_2}\left\{\tilde{\boldsymbol{K}}_2^{*'}(\eta_2;\boldsymbol{\pi}^0,\lambda_1^0)\left(\boldsymbol{H}_2^*(\eta_2)\otimes\boldsymbol{V}^*\right)\tilde{\boldsymbol{K}}_2^*(\eta_2;\boldsymbol{\pi}^0,\lambda_1^0)\right\},$$

where $\boldsymbol{V}^*$ *is as in Theorem 2, and* $\boldsymbol{H}_j^*(\eta_j)$*,* $j = 1, 2$*, is as in Theorem 3.*

Theorem 4 characterizes the limiting distribution of sup-$Wald_T(2|1)$ for any number of breaks $h = 0, 1, 2, \ldots, h_{\max}$ in RF, but the specific form of the distribution depends on h and $\boldsymbol{\pi}$.

As for Theorems 1 and 2, the distributions in Theorems 3 and 4 are the same as in BP for exogenous regressors or for a stable RF, and for an unstable RF the distributions depend on the number and location of the RF breaks.

5 Bootstrap

In this section, we introduce the bootstrap analogues of the sup-F and sup-$Wald$ statistics for multiple breaks. In particular, we introduce the bootstrapped versions of (10) and (11) for the hypothesis in (a) and the bootstrap analogues of the statistics (16) and (18) for the hypothesis in (b); these hypothesis are defined just after (5).

For the case when the error terms u_t and $\boldsymbol{v}_t$ are IID but potentially contemporaneously correlated (and the regressors are exogenous or endogenous), we show that the IR (IID recursive) bootstrap and the IF (IID fixed-regressor) bootstrap based on the bootstrap analogues of the sup-F tests in (10) and (16), are asymptotically

valid. The ideas behind the IR and IF bootstraps are introduced by Efron (1979). In each case, the bootstrap samples are obtained by re-sampling from the (joint) empirical distribution of the (centered) residuals. The IR bootstrap computes the bootstrapped y_t, $\boldsymbol{x}_t$ recursively from the SE and RF, while the IF bootstrap computes them by adding the estimated conditional mean to the bootstrap residuals, while keeping the lags of y_t, $\boldsymbol{x}_t$ present in $\boldsymbol{z}_t$ as fixed and equal to the sample values. The validity of these bootstraps for structural change tests and unknown breaks has not been previously established even for models with exogenous regressors, although these bootstraps have been used extensively in many papers.[16] For example, Christiano (1992) has used the IR bootstrap for testing for a break in the U.S. GNP. Diebold and Chen (1996) provide simulation results for testing for a single break in stationary autoregressive models using the IR bootstrap and conclude that the bootstrap is preferred to the asymptotic test proposed by Andrews (1993). Clark (2006) confirms the conclusions of Diebold and Chen (1996) by considering a larger diversity of data generating processes drawn from 1984-2002 estimates of autoregressive models used for modelling inflation. In addition, Antoshin, Berg, and Souto (2008) show in Monte Carlo simulations that the IR bootstrap is preferred to the asymptotic tests for multiple breaks proposed by Bai and Perron (1998), and Berg, Ostry, and Zettelmeyer (2012) use the IR bootstrap to show evidence of structural breaks in the economic growth of 140 countries. Moreover, Banerjee, Lazarova, and Urga (2002), de Peretti and Urga (2004) and Bergamelli and Urga (2013) show through Monte Carlo simulations that the IR bootstrap detects multiple breaks in systems of equations.

The IID assumption of the disturbances does not always follow from economic models and the conditional (and unconditional) heteroskedasticity in the residuals is more frequent in many estimated dynamic regression models in finance and macroeconomics; see for example Bollerslev (1986), Hodrick (1992) and Bekaert and Hodrick (2001). However, these papers impose a parametric structure for the conditional heteroskedasticity. In addition, for multivariate conditional heteroskedastic models, it is difficult to obtain reliable numerical estimates of conditional covariance without imposing additional restrictions on the covariance structure of the disturbances; see among others Bollerslev (1990) and Ledoit, Santa-Clara, and Wolf (2003). In contrast, the wild bootstrap allows for a nonparametric treatment of the conditional (and unconditional) heteroskedasticity, thus avoiding the difficulties mentioned above. 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. We consider both the wild recursive (WR) bootstrap and the wild fixed (WF) version of the wild bootstrap in dynamic models that takes into account the possibility of the errors u_t and $\boldsymbol{v}_t$ being conditionally and unconditionally heteroskedastic, as well as contemporaneously correlated. The idea of the wild bootstrap is to replicate heteroskedasticity by computing the bootstrap residuals as the product of the initial residual on a particular observation multiplied by a random variable independent of the data, with mean 0 and variance 1. There are several choices for the distribution of the random variable (auxiliary distribution): Gonçalves and Kilian (2004) use the standard normal distribution, while Mammen (1993) and Liu (1988) suggested a two-point distribution. In this paper, we report simulation results for Liu's two-point distribution, which performed the best compared to the other distributions. This conclusion is similar to Davidson and Flachaire (2008) and Davidson and MacKinnon (2010).

With unknown break location, the asymptotic validity of the WR and WF bootstraps for structural change tests has not been shown previously, although the WR bootstrap has already been used in practice. For example,

[16] Hansen (2000) proves validity of a different IF bootstrap for a structural change test and exogenous regressors; see comments on the next page.

O'Reilly and Whelan (2004) show by simulations that the WR bootstrap works well across a wide variety of data generating processes used in macroeconomics. Building on this conclusion, Levin and Piger (2006) use the WR bootstrap to find evidence of structural breaks in the intercept of autoregressive models for inflation for eight OECD countries.

In this paper, we show that the WR and WF bootstraps based on the bootstrap analogues of the *sup-Wald* tests in (11) and (18) are asymptotically valid. As we will see next, the WR bootstrap requires more stringent assumptions than the WF bootstrap on u_t and $\boldsymbol{v}_t$, in order to ensure convergence to the correct covariance matrix. The assumptions that guarantee the validity of the WR bootstrap allow, for example, for the regression disturbances to be from (G)ARCH and stochastic volatility models with symmetric errors and finite fourth moment, while the WF bootstrap covers also the (G)ARCH and stochastic volatility models with asymmetric errors and finite fourth moments; see Gonçalves and Kilian (2004). Moreover, in contrast with Gonçalves and Kilian (2004), we allow for both conditional and unconditional heteroskedasticity (of unknown form) provided a global homoskedasticity assumption is satisfied (see Assumptions 13(ii) and 14). The global homoskedasticity condition is satisfied for example by models with seasonal heteroskedasticity; see Cavaliere, Rahbek, and Taylor (2010). The global homoskedasticity assumption is also satisfied when there are shrinking breaks in the unconditional variance of the error term. Finally, our WR bootstrap for structural changes relies on the existence of moments slightly larger than four, as opposed to Gonçalves and Kilian (2004) who assume the existence of moments of order eight in the context of bootstrapping the coefficients of autoregressive models.[17]

A different version of the IF and WF bootstraps was proposed by Hansen (2000) in the context of structural breaks in dynamic models estimated by OLS when there are fixed breaks in the marginal distribution of the regressors. In this paper, we do not allow for fixed breaks in the marginal distribution of the regressors but we allow for shrinking breaks in the marginal distribution of the regressors.[18] We show that the IF and WF bootstraps, that simply add the (IID and wild) bootstrap residuals to the conditional mean, as in Kreiss (1997) and Gonçalves and Kilian (2004), are asymptotically valid for structural change tests in dynamic models estimated by OLS or 2SLS.[19]

In Section 5.1 we introduce the four bootstrap methods and the assumptions for their validity. We consider the IID recursive bootstrap (IR bootstrap) (Section 5.1.1); the wild recursive bootstrap (WR bootstrap) (Section 5.1.2); the IID fixed-regressor bootstrap (IF bootstrap) (Section 5.1.3) and the wild fixed-regressor bootstrap (WF bootstrap) (Section 5.1.4). In Section 5.2 we present the theorems that show the asymptotic validity of all these bootstrap methods. In Section 6 and Appendix C we provide the results of a simulation study for all these bootstrap tests and compare them to the asymptotic sup-F and sup-$Wald$ tests based on the distributions derived in BP, in HHB, and in Theorems 1-4 above.

5.1 Methods and assumptions

For each of the IR, IF, WR and WF bootstraps, we describe in detail how to generate the bootstrap disturbances and the bootstrap samples. Once a bootstrap sample is generated, all the test statistics, parameter estimates,

[17] The difference stems from the fact that Gonçalves and Kilian (2004) use a Functional Central Limit Theorem (FCLT) that requires the convergence of sample moments while we use the FCLT of Wooldridge and White (1988), Theorem 2.11 which requires only verifying the convergence of the population moments.

[18] We can allow for fixed (large) breaks instead of shrinking (moderate) breaks. Our simulations show that the bootstrap works well in both cases, but our theory is developed under shrinking breaks because the notation and the proofs are greatly simplified.

[19] Hansen's fixed-regressor bootstrap considers as bootstrap sample the (IID or wild) bootstrap residuals, while here a bootstrap sample is obtained by adding the (IID or wild) bootstrap residuals to the estimated conditional mean (see Sections 5.1.3 and 5.1.4).

sum of squared residuals that are calculated exactly as in the original sample, but on a bootstrap sample, are referred to as *bootstrap analogues* and denoted with a b superscript. Only when this calculation is unclear, we elaborate on it further.

5.1.1 IID recursive bootstrap (IR bootstrap)

(a) The null hypothesis $H_0 : m = 0$ against the alternative hypothesis $H_1 : m = k$.

Given the parameter estimates from RF which we denote by $\hat{\boldsymbol{\Delta}}(t,T) = \sum_{j=1}^{\hat{h}+1} 1_{t\in[\hat{T}^*_{j-1}+1,\hat{T}^*_j]}\,\hat{\boldsymbol{\Delta}}_{(j)}$, with $\hat{h} \geq 0$ estimated breaks, and the parameter estimates from SE under the null hypothesis $m = 0$, which we denote by $\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1}$, an IR bootstrap sample is obtained recursively by adding to the estimated conditional mean the bootstrap residuals obtained by resampling from the joint empirical distribution function of the centered 2SLS residuals, i.e.

$$\begin{aligned} &\boldsymbol{x}_t^{b\prime} = \boldsymbol{z}_t^{b\prime}\hat{\boldsymbol{\Delta}}(t,T) + \boldsymbol{v}_t^{b\prime}, \qquad y_t^b = \boldsymbol{x}_t^{b\prime}\hat{\boldsymbol{\beta}}_{\boldsymbol{x}} + \boldsymbol{z}_{1,t}^b\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1} + u_t^b = \boldsymbol{w}_t^{b\prime}\hat{\boldsymbol{\beta}} + u_t^b, \\ &\boldsymbol{z}_t^b = \left(\boldsymbol{r}_t', y_{t-1}^b, \boldsymbol{x}_{t-1}^{b\prime}, y_{t-2}^b, \boldsymbol{x}_{t-2}^{b\prime}, \ldots, \boldsymbol{x}_{t-\tilde{q}_2}^{b\prime}\right)', \boldsymbol{z}_{1,t}^b = \left(\boldsymbol{r}_{1,t}', y_{t-1}^b, \boldsymbol{x}_{t-1}^{b\prime}, y_{t-2}^b, \boldsymbol{x}_{t-2}^{b\prime} \ldots, \boldsymbol{x}_{t-\tilde{q}_1}^{b\prime}\right)', \boldsymbol{w}_t^b = \left(\boldsymbol{x}_t^{b\prime}, \boldsymbol{z}_{1,t}^{b\prime}\right)', \end{aligned} \tag{33}$$

and the bootstrap residuals u_t^b and $\boldsymbol{v}_t^b$, $t = 1, \ldots, T$, are independently drawn with replacement from the joint empirical distribution of the centered residuals $(\hat{u}_t - \bar{u}, (\hat{\boldsymbol{v}}_t - \bar{\boldsymbol{v}})')'$, with $\hat{u}_t = y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_{\boldsymbol{x}} - \boldsymbol{z}_{1,t}'\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1}$, $\bar{u} = T^{-1}\sum_{t=1}^T \hat{u}_t$, $\hat{\boldsymbol{v}}_t' = \boldsymbol{x}_t' - \boldsymbol{z}_t'\hat{\boldsymbol{\Delta}}(t,T)$ and $\bar{\boldsymbol{v}} = T^{-1}\sum_{t=1}^T \hat{\boldsymbol{v}}_t$. Drawing the bootstrap residuals from the joint distribution of the residuals preserves the contemporaneous correlation between u_t and $\boldsymbol{v}_t$. The residuals need not be re-scaled before bootstrapping since sup-F and sup-$Wald$ are scale invariant.[20] Note that for constructing $\boldsymbol{z}_t^b$ and $\boldsymbol{z}_{1t}^b$, in all bootstraps, we use the true number of lags in the RF and SE, if known, that is, $\tilde{p}_2, \tilde{p}_1$ lags of y_t and $\tilde{q}_2, \tilde{q}_1$ lags of $\boldsymbol{x}_t$. Because $\boldsymbol{z}_t$ contains $\boldsymbol{z}_{1,t}$, $\tilde{p}_2 \geq \tilde{p}_1$ and $\tilde{q}_2 \geq \tilde{q}_1$. The starting values for the recursions in (33) were set equal to the first observations in the original sample, that is $y_s^b = y_s$, $s = 1, \ldots, \tilde{p}_2$, and $\boldsymbol{x}_s^{b\prime} = \boldsymbol{x}_s'$, $s = 1, \ldots, \tilde{q}_2$, in line with Davidson and MacKinnon (1993), but they can also be drawn at random from the full original sample (see Berkowitz and Kilian (2000)).

The IR bootstrap analogue of (10) for testing $H_0 : m = 0$ against $H_1 : m = k$ is

$$\sup\text{-}F_T^b = \sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}_\epsilon} F_T^b(\boldsymbol{\lambda}), \qquad F_T^b(\boldsymbol{\lambda}) = \left(\frac{T-(k+1)p}{kp}\right)\left(\frac{SSR_0^b - SSR_k^b(\boldsymbol{\lambda};\,\hat{\boldsymbol{\beta}}^b(\boldsymbol{\lambda}))}{SSR_k^b(\boldsymbol{\lambda};\,\hat{\boldsymbol{\beta}}^b(\boldsymbol{\lambda}))}\right), \tag{34}$$

where $F_T^b(\boldsymbol{\lambda})$ is the bootstrap analogue of $F_T(\boldsymbol{\lambda})$ defined in (9), using SSR_0^b and $SSR_k^b(\boldsymbol{\lambda};\,\hat{\boldsymbol{\beta}}^b(\boldsymbol{\lambda}))$ as the bootstrap analogues of SSR_0 and $SSR_k(\boldsymbol{\lambda};\,\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}))$. In computing these bootstrap analogues, the regressors generated from the first-stage use the $\hat{h}$-break fraction partition $\hat{\boldsymbol{\pi}}$ previously estimated on the original sample. So the RF break fraction estimates are not bootstrapped, because they are converging fast enough so that their randomness does not affect the results. More exactly, we use $\hat{\boldsymbol{x}}_t^{b\prime} = \boldsymbol{z}_t^{b\prime}\hat{\boldsymbol{\Delta}}^b(t,T)$, $\hat{\boldsymbol{w}}_t^b = \left(\hat{\boldsymbol{x}}_t^{b\prime}, \boldsymbol{z}_{1,t}^{b\prime}\right)'$, and we let $\hat{\boldsymbol{\Delta}}^b(t,T)$ be the bootstrap analogue of $\hat{\boldsymbol{\Delta}}(t,T)$, i.e.

$$\hat{\boldsymbol{\Delta}}^b(t,T) = \sum_{j=1}^{\hat{h}+1} 1_{t\in[\hat{T}^*_{j-1}+1,\hat{T}^*_j]}\hat{\boldsymbol{\Delta}}^b_{(j)}, \qquad \hat{\boldsymbol{\Delta}}^b_{(j)} = \left(\sum_{t=\hat{T}^*_{j-1}+1}^{\hat{T}^*_j} \boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}\right)^{-1}\left(\sum_{t=\hat{T}^*_{j-1}+1}^{\hat{T}^*_j} \boldsymbol{z}_t^b\boldsymbol{x}_t^{b\prime}\right). \tag{35}$$

To show the asymptotic validity of the IR bootstrap based on (34), we impose the following assumption.

Assumption IR(a). *Assumptions 1, 4-12 hold. In addition,*

[20] See for example: Flachaire (1999), Davidson (2007), Davidson and MacKinnon (2010).

(i) $\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}_{t-i}') = 0$ *for all* t, $i \neq j$, $i \geq 1$, $j \geq 1$.

(ii) $\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-i}') = \boldsymbol{\Omega}^2$ *is uniformly bounded for all* t, $i \geq 0$.

(iii) y_t^b *and* $\boldsymbol{x}_t^b$ *are obtained as in (33) with the bootstrap residuals such that* $u_t^b = \hat{u}_{\varsigma_t} - \bar{u}$, $\boldsymbol{v}_t^b = \hat{\boldsymbol{v}}_{\varsigma_t} - \bar{\boldsymbol{v}}$ *with* $\varsigma_t \overset{IID}{\sim} U(1,T)$, $t = 1,\ldots,T$.

Assumption IR(a)(i),(ii) are homogeneity assumptions on the cross-moments of $\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-i}$ and $\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-j}$, implied by Assumptions 9 and 12, but stated here nevertheless, for clarity and ease of comparison to the corresponding Assumption WR(a). For example, Assumption IR(a)(i) is implied by the martingale diffference Assumption 12(i), but it is more general. Assumption IR(a)(ii) does not allow $\boldsymbol{\varepsilon}_t$ to follow a GARCH model[21], or to exhibit (shrinking) variance breaks. These restrictions make sense as the IID bootstrap is not suited to handle heteroskedasticity. Assumption IR(a)(iii) states that the bootstrap residuals are drawn independently with replacement from the joint distribution of the centered residuals, and so they can be written as $u_t^b = \hat{u}_{\varsigma_t} - \bar{u}$ and $\boldsymbol{v}_t^b = \hat{\boldsymbol{v}}_{\varsigma_t} - \bar{\boldsymbol{v}}$ with $\varsigma_t \overset{IID}{\sim} U(1,T)$.

(b) The null hypothesis $H_0 : m = 1$ against the alternative hypothesis $H_1 : m = 2$.

To test the null of one SE break against one additional SE break, we now consider the bootstrap analogue of (16). Under the null hypothesis of one SE break, estimated at $\hat{T}_1 = [T\hat{\lambda}_1]$, the bootstrap samples are generated recursively as follows:

$$\boldsymbol{x}_t^{b'} = \boldsymbol{z}_t'\hat{\boldsymbol{\Delta}}(t,T) + \boldsymbol{v}_t^{b'}, \tag{36}$$

$$y_t^b = 1_{t\leq\hat{T}_1}\left(\boldsymbol{x}_t^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(1)} + \boldsymbol{z}_{1,t}^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(1)} + u_{t,(1)}^b\right) + 1_{t>\hat{T}_1}\left(\boldsymbol{x}_t^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(2)} + \boldsymbol{z}_{1,t}^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(2)} + u_{t,(2)}^b\right) = \boldsymbol{w}_t^{b\prime}\hat{\boldsymbol{\beta}}(t,T) + u_t^b, \tag{37}$$

with $\hat{\boldsymbol{\beta}}(t,T) = \hat{\boldsymbol{\beta}}_{(1)}1_{t\leq\hat{T}_1} + \hat{\boldsymbol{\beta}}_{(2)}1_{t>\hat{T}_1}$, $\hat{\boldsymbol{\beta}}_{(1)} = \left(\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(1)}', \hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(1)}'\right)'$ and $\hat{\boldsymbol{\beta}}_{(2)} = \left(\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(2)}', \hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(2)}'\right)'$ the 2SLS estimators in samples $[1,\hat{T}_1]$, respectively $[\hat{T}_1+1,T]$, $u_t^b = 1_{t\leq\hat{T}_1}u_{t,(1)}^b + 1_{t>\hat{T}_1}u_{t,(2)}^b$, $\boldsymbol{v}_t^{b\prime} = 1_{t\leq\hat{T}_1}\boldsymbol{v}_{t,(1)}^{b'} + 1_{t>\hat{T}_1}\boldsymbol{v}_{t,(2)}^{b'}$, where $\left(u_{t,(i)}^b, \boldsymbol{v}_{t,(i)}^{b'}\right)'$, $i = 1,2$, are independently drawn with replacement from the joint distribution of the centered restricted residuals $\left(\hat{u}_{t,(i)} - \bar{u}_{(i)}, \hat{\boldsymbol{v}}_{t,(i)}' - \bar{\boldsymbol{v}}_{(i)}'\right)'$,

$$\hat{u}_{t,(1)} = y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(1)} - \boldsymbol{z}_{1,t}'\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(1)}, \quad t = 1,\ldots,\hat{T}_1, \tag{38}$$

$$\hat{u}_{t,(2)} = y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(2)} - z_{1,t}'\hat{\beta}_{\boldsymbol{z}_1,(2)}, \quad t = \hat{T}_1+1,\ldots,T, \tag{39}$$

$\bar{u}_1 = \hat{T}_1^{-1}\sum_{t=1}^{\hat{T}_1}\hat{u}_{t,(1)}$, $\bar{u}_2 = (T-\hat{T}_1)^{-1}\sum_{t=\hat{T}_1+1}^{T}\hat{u}_{t,(2)}$, and $\hat{\boldsymbol{v}}_t' = \boldsymbol{x}_t' - \boldsymbol{z}_t'\hat{\boldsymbol{\Delta}}(t,T)$, $\hat{\boldsymbol{v}}_{t,(1)} = \hat{\boldsymbol{v}}_t 1_{t\leq\hat{T}_1}$, $\hat{\boldsymbol{v}}_{t,(2)} = \hat{\boldsymbol{v}}_t 1_{t>\hat{T}_1}$, where $\bar{\boldsymbol{v}}_{(1)}$ and $\bar{\boldsymbol{v}}_{(2)}$ are their sample means ($p_1 \times 1$ vectors), respectively.

In the bootstrap samples, we do not re-estimate $\hat{T}_1 = [T\hat{\lambda}_1]$, we take it as given, because $\hat{\lambda}_1$ it converges fast enough to λ_1^0 so that its randomness can be ignored. Note that the bootstrap samples in (36)-(37) are generated under the null hypothesis of $m = 1$ breaks in SE. [22]

The IR bootstrap analogue of (16) is:

$$\sup\text{-}F_T^b(2|1) = \max\left(\frac{SSR_1^b(\hat{T}_1) - SSR_1^b(\hat{\tau}_1|\hat{T}_1)}{SSR_1^b(\hat{T}_1)/(\hat{T}_1 - p)}, \frac{SSR_2^b(\hat{T}_1) - SSR_2^b(\hat{\tau}_2|\hat{T}_1)}{SSR_2^b(\hat{T}_1)/(T - \hat{T}_1 - p)}\right), \tag{40}$$

where $SSR_i^b(\hat{T}_1)$ and $SSR_i^b(\tau_i|\hat{T}_1)$ are the bootstrap analogues of $SSR_i(\hat{T}_1)$ and $SSR_i(\tau_i|\hat{T}_1)$ for $i = 1,2$, computed with the break $\hat{T}_1$ in the SE and the $\hat{h}$-partition $\hat{\boldsymbol{\pi}}$ in the RF.

To show the asymptotic validity of the IR bootstrap based on (40), we impose the following assumption.

[21] Suppose the regressors are exogenous, so that $\boldsymbol{\varepsilon}_t = u_t$. Also let u_t follow a GARCH process. Then u_t^2 is an ARMA process, so $\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-i}') = \mathrm{E}(u_t^2u_{t-i}^2)$ will depend on i in general.

[22] We can generalise (37) to $m > 1$ breaks in SE, in which case $\hat{\boldsymbol{\beta}}(t,T) = \sum_{i=1}^{m+1}\hat{\boldsymbol{\beta}}_{(i)}$, where $\hat{\boldsymbol{\beta}}_{(i)}$ is the estimate for the sub-sample $[\hat{T}_{i-1}+1, \hat{T}_i]$.

Assumption IR(b). *Assumptions 2-3 and Assumption IR(a)(i)-(ii) hold. Additionally,*

(iii) y_t^b *and* $\boldsymbol{x}_t^b$ *are obtained as in (36) and (37) with the bootstrap residuals such that* $u_{t,(i)}^b = \hat{u}_{\varsigma_t,(i)} - \bar{u}_{(i)}$, $\boldsymbol{v}_{t,(i)}^b = \hat{\boldsymbol{v}}_{\varsigma_t,(i)} - \bar{\boldsymbol{v}}_{(i)}$, $i = 1, 2$, *with* $\varsigma_{t,(1)} \overset{IID}{\sim} U(1, \hat{T}_1)$, $\varsigma_{t,(2)} \overset{IID}{\sim} U(1, T - \hat{T}_1)$.

Assumption IR(b)(iii) highlights the fact that bootstrap residuals are drawn independently and with replacement from the joint distribution of the centered 2SLS residuals of each of the two subsamples, i.e. $\{\hat{u}_t - \bar{u}_{(1)}, \hat{\boldsymbol{v}}_t - \bar{\boldsymbol{v}}_{(1)}\}_{t=1}^{\hat{T}_1}$ and $\{\hat{u}_t - \bar{u}_{(2)}, \hat{\boldsymbol{v}}_t - \bar{\boldsymbol{v}}_{(2)}\}_{t=\hat{T}_1+1}^{T}$, so that they can be written $u_{t,(i)}^b = \hat{u}_{\varsigma_{t,(i)}} - \bar{u}_{(i)}$, $\boldsymbol{v}_{t,(i)}^b = \hat{\boldsymbol{v}}_{\varsigma_{t,(i)}} - \bar{\boldsymbol{v}}_{(i)}$, $i = 1, 2$, where $\varsigma_{t,(1)} \overset{IID}{\sim} U(1, \hat{T}_1)$, $t = 1, \ldots, \hat{T}_1$ and $\varsigma_{t,(2)} \overset{IID}{\sim} U(1, T - \hat{T}_1)$, $t = 1, \ldots, (T - \hat{T}_1)$.

5.1.2 Wild recursive bootstrap (WR bootstrap)

(a) The null hypothesis $H_0 : m = 0$ against the alternative hypothesis $H_1 : m = k$.

The bootstrap samples are generated as for the IR(a) bootstrap, recursively as in (33) under the null hypothesis that there are $m = 0$ breaks in SE, except that the bootstrap residuals are obtained as $u_t^b = \hat{u}_t \varsigma_t$ and $\boldsymbol{v}_t^b = \hat{\boldsymbol{v}}_t \varsigma_t$, where $\hat{u}_t$ and $\hat{\boldsymbol{v}}_t$ are the (non-centered) residuals under the null hypothesis, $\varsigma_t \overset{IID}{\sim} (0, 1)$ and they are independent of $\hat{u}_t$ and $\hat{\boldsymbol{v}}_t$. For the WR bootstrap, the residuals need not to be centered because ς_t has mean zero, so u_t^b and $\boldsymbol{v}_t^b$ have mean zero.

The recursive wild bootstrap analogue of the sup-$Wald$ statistic from (11) is

$$\sup\text{-}Wald_T^b = \sup_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}_\epsilon} Wald_T^b(\boldsymbol{\lambda}), \tag{41}$$

with $Wald_T^b(\boldsymbol{\lambda}) = T\hat{\boldsymbol{\beta}}^b(\boldsymbol{\lambda})' \boldsymbol{R}_k' \left(\boldsymbol{R}_k \hat{\boldsymbol{V}}^b(\boldsymbol{\lambda}) \boldsymbol{R}_k'\right)^{-1} \boldsymbol{R}_k \hat{\boldsymbol{\beta}}^b(\boldsymbol{\lambda})$ the bootstrap analogue of $Wald_T(\boldsymbol{\lambda})$; $\hat{\boldsymbol{\beta}}^b(\boldsymbol{\lambda})$ the bootstrap analogue of $\hat{\boldsymbol{\beta}}(\boldsymbol{\lambda})$ based on the k-partition $\boldsymbol{\lambda}$ under the alternative of k breaks;

$$\hat{\boldsymbol{V}}^b(\boldsymbol{\lambda}) = \operatorname{diag}\left(\hat{\boldsymbol{V}}_{(1)}^b, \ldots, \hat{\boldsymbol{V}}_{(k)}^b\right); \qquad \hat{\boldsymbol{Q}}_{(i)}^b = T^{-1} \sum_{t=T_{i-1}+1}^{T_i} \hat{\boldsymbol{w}}_t^b \hat{\boldsymbol{w}}_t^{b\prime}; \qquad \hat{\boldsymbol{V}}_{(i)}^b = \left(\hat{\boldsymbol{Q}}_{(i)}^b\right)^{-1} \hat{\boldsymbol{M}}_{(i)}^b \left(\hat{\boldsymbol{Q}}_{(i)}^b\right)^{-1}; \tag{42}$$

$\hat{\boldsymbol{M}}_{(i)}^b = T^{-1} \sum_{t=T_{i-1}+1}^{T_i} \hat{\boldsymbol{\Upsilon}}^{b\prime}(t, T) \boldsymbol{z}_t^b \left(u_t^b + \boldsymbol{v}_t^{b\prime} \hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^b\right) \left(\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^{b\prime} \boldsymbol{v}_t^b + u_t^b\right) \boldsymbol{z}_t^{b\prime} \hat{\boldsymbol{\Upsilon}}^b(t, T)$ the bootstrap analogue of $\hat{\boldsymbol{M}}_{(i)}$ from (14) for $i = 1, \ldots, k+1$; $\hat{\boldsymbol{\Upsilon}}^b(t, T) = (\hat{\boldsymbol{\Delta}}^b(t, T), \boldsymbol{\Pi})$; $\hat{\boldsymbol{\Delta}}^b(t, T)$ the bootstrap analogue of $\hat{\boldsymbol{\Delta}}(t, T)$ computed using the $\hat{h}$-partition $\hat{\boldsymbol{\pi}}$ which was estimated on the original sample.

To show the validity of the WR bootstrap based on (41), we need to restrict Assumption 13(ii) and impose additional assumptions.

Assumption 14. *(i) Assumption 12(i) holds. In addition, for all $l > 0$,*

(ii) $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{[Ts]} \mathrm{E}(\boldsymbol{\varepsilon}_{t-l} \boldsymbol{\varepsilon}_{t-l}') = s\boldsymbol{\Omega}$ *uniformly in s;*

(iii) $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{[Ts]} \mathrm{E}(\boldsymbol{\varepsilon}_{t-l} \boldsymbol{\varepsilon}_{t-l}' \mid \mathcal{F}_{t-l}) = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{[Ts]} \boldsymbol{\Omega}_{t-l} = s\boldsymbol{\Omega}$ *in probability uniformly in s. Moreover, we have that* $\mathrm{E} \left\| \boldsymbol{\varepsilon}_{t-l} \boldsymbol{\varepsilon}_{t-l}' - \mathrm{E}(\boldsymbol{\varepsilon}_{t-l} \boldsymbol{\varepsilon}_{t-l}' \mid \mathcal{F}_{t-l}) \right\|^{\bar{d}} < \infty$, *for some $\bar{d} > 1$.*

Both Assumptions 13 and 14 impose a global homoskedasticity condition; see Davidson (1994) pp.454-455 and Cavaliere, Rahbek, and Taylor (2010) p.1723 and Note 1. They imply that $\mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$, $\mathrm{E}((\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') \otimes (\boldsymbol{z}_t \boldsymbol{z}_t'))$ and $\mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mid \mathcal{F}_t)$ can change over time provided they are asymptotically stable.[23] Assumptions 13(ii) and 14 allow, for example, for models with seasonal heteroskedasticity (see Cavaliere, Rahbek, and Taylor (2010)), shrinking

[23] We mean that $1/(T(s' - s)) \sum_{t=[Ts]+1}^{[Ts']} \mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') \to \boldsymbol{\Omega}$, $1/(T(s' - s)) \sum_{t=[Ts]+1}^{[Ts']} \mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mid \mathcal{F}_t) \xrightarrow{p} \boldsymbol{\Omega}$, $1/(T(s' - s)) \sum_{t=[Ts]+1}^{[Ts']} \mathrm{E}((\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') \otimes (\boldsymbol{z}_t \boldsymbol{z}_t')) \to \tilde{\boldsymbol{\Sigma}}$, for all $s' < s \in [0, 1]$.

breaks in the conditional and unconditional variance of the error term $\boldsymbol{\varepsilon}_t$, and shrinking breaks in the marginal distribution of the instruments $\boldsymbol{z}_t$. Assumption 14(ii) is weaker than Assumption A(ii) in Gonçalves and Kilian (2004) which requires covariance stationary error terms.

Assumption WR(a). *Assumptions 1, 4-11, and 14 hold. In addition, Assumption 13(ii) holds with:*

(i) $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{[Ts]} \mathrm{E}((\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')\otimes(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-j}')) = 0$ *uniformly in* s, $i \neq j$, *for all* t, $i \geq 1$, $j \geq 1$;

(ii) $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{[Ts]} \mathrm{E}((\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')\otimes(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-i}')) = s\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon},ii}$ *uniformly in* s, *where* $\left\|\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon},ii}\right\| < \infty$, *for all* t, $i \geq 0$. *When* $i = 0$, $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon},00} = \boldsymbol{\Omega}\otimes\boldsymbol{\Omega}$;

(iii) $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{[Ts]} \mathrm{E}((\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')\otimes(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{r}_{t-j}')) = 0$ *uniformly in* s, *for all* t, $i \geq 1$, $j \geq 0$.

Also, we have the following conditions:

(iv) $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{[Ts]} \mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' \mid \mathcal{F}_t)\otimes\left(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-i}'\right) = s\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon},ii}$ *in probability uniformly in* s, *for all* $i \geq 1$;

(v) $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{[Ts]} \mathrm{E}(\boldsymbol{\varepsilon}_{t-l}\boldsymbol{\varepsilon}_{t-l}' \mid \mathcal{F}_{t-l})\otimes\left(\boldsymbol{r}_{t-i}\boldsymbol{r}_{t-j}'\right) = s\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{r},ij}$ *in probability uniformly in* s, *for all* $i,j,l \geq 0$;

(vi) y_t^b *and* $\boldsymbol{x}_t^b$ *are obtained as in (33) with the bootstrap residuals such that* $u_t^b = \hat{u}_t\varsigma_t$ *and* $\boldsymbol{v}_t^b = \hat{\boldsymbol{v}}_t\varsigma_t$ *where* $\varsigma_t \overset{IID}{\sim} (0,1)$, $t = 1,\ldots,T$; $\mathrm{E}^b|\varsigma_t|^{4+\epsilon} = \bar{c} < \infty$, *for some* $\epsilon > 0$, *for all* t.

These are further homogeneity conditions that facilitate the WR bootstrap validity proofs, in the absence of the martingale difference and conditional homoskedasticity Assumption 12 on $\boldsymbol{\varepsilon}_t$. To understand Assumption WR(a)(i)-(iii), examine Assumption 13 and notice that it allows $\mathrm{E}((\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')\otimes(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-j}')) \neq 0$ for $i \neq j$. However, by the WR bootstrap design, $\mathrm{E}^b((\boldsymbol{\varepsilon}_t^b\boldsymbol{\varepsilon}_t^{b'})\otimes(\boldsymbol{\varepsilon}_{t-i}^b\boldsymbol{\varepsilon}_{t-j}^{b'})) = (\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t')\otimes(\hat{\boldsymbol{\varepsilon}}_{t-i}\hat{\boldsymbol{\varepsilon}}_{t-j}')\,\mathrm{E}^b(\varsigma_t^2\varsigma_{t-i}\varsigma_{t-j}) = (\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t')\otimes(\hat{\boldsymbol{\varepsilon}}_{t-i}\hat{\boldsymbol{\varepsilon}}_{t-j}')\,\mathrm{E}^b(\varsigma_t^2)\,\mathrm{E}^b(\varsigma_{t-i})\,\mathrm{E}^b(\varsigma_{t-j}) = 0$ by Assumption WR(a)(vi) if $i \neq j$ since $\mathrm{E}^b(\varsigma_t) = 0$ for all t. In the WR bootstrap proofs, we need partial sums of both these quantities to converge to the same number; therefore, we impose Assumption WR(a)(i). By similar reasoning, we impose Assumption WR(a)(iii) to mimic the condition $\mathrm{E}^b((\boldsymbol{\varepsilon}_t^b\boldsymbol{\varepsilon}_t^{b'})\otimes(\boldsymbol{\varepsilon}_{t-i}^b\boldsymbol{r}_{t-j}')) = 0$ for $\mathrm{E}((\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')\otimes(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{r}_{t-j}'))$ in the limit, at all t, $i \geq 1$, $j \geq 0$.

Denote by $s\tilde{\tilde{\boldsymbol{\Sigma}}} = s\begin{pmatrix} \tilde{\tilde{\boldsymbol{\Sigma}}}_u & \tilde{\tilde{\boldsymbol{\Sigma}}}_{u\boldsymbol{v}} \\ \tilde{\tilde{\boldsymbol{\Sigma}}}_{u\boldsymbol{v}}' & \tilde{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{v}} \end{pmatrix} = \lim_{T\to\infty} \boldsymbol{V}_T(s)$ when Assumption 13 restricted by Assumption WR(a)(i)-(iii) is satisfied.

Assumption WR(a)(ii) and Assumption WR(a)(iv) can be interpreted as global homoskedasticity assumptions. Unlike Assumption IR(a)(ii), Assumption WR(a)(ii) allows $\boldsymbol{\varepsilon}_t$ to follow (G)ARCH models and stochastic volatility models with finite fourth moments and symmetric errors, and to exhibit shrinking variance breaks. Assumption WR(a)(v) is just a regularity condition, and Assumption WR(a)(vi) states that the bootstrap residuals are obtained by multiplying the actual residuals by a random variable, independent of the data, with mean 0 and variance 1.

(b) The null hypothesis $H_0 : m = 1$ against the alternative hypothesis $H_1 : m = 2$.

The bootstrap samples are generated as for the IR bootstrap case (b), recursively as in (36)-(37) under the null hypothesis that there are $m = 1$ breaks in SE, except that the bootstrap residuals are $u_{t,(i)}^b = \hat{u}_{t,(i)}\varsigma_{t,(i)}$ and $\boldsymbol{v}_{t,(i)}^b = \hat{\boldsymbol{v}}_{t,(i)}\varsigma_{t,(i)}$, $i = 1,2$, where $\varsigma_{t,(i)} \overset{IID}{\sim} (0,1)$.

The WR bootstrap analogue of (18) is

$$Wald_T^b(2|1) = \max\left(Wald_{T,1}^b(2|1;\hat{T}_1),\ Wald_{T,2}^b(2|1;\hat{T}_1)\right), \tag{43}$$

where $Wald^b_{T,1}(2|1;\hat{T}_1)$ and $Wald^b_{T,2}(2|1;\hat{T}_1)$ are the bootstrap sup-$Wald$ statistics for testing for 0 breaks against one break in the sub-sample before $\hat{T}_1$ and after $\hat{T}_1$ respectively:

$$Wald^b_{T,1}(2|1;\hat{T}_1) = \sup_{\tau_1 \in \boldsymbol{\Lambda}_1} T\hat{\boldsymbol{\beta}}^b(\eta_1)' \boldsymbol{R}_1' [\boldsymbol{R}_1 \hat{\boldsymbol{V}}^b(\eta_1) \boldsymbol{R}_1']^{-1} \boldsymbol{R}_1 \hat{\boldsymbol{\beta}}^b(\eta_1), \tag{44}$$

$$Wald^b_{T,2}(2|1;\hat{T}_1) = \sup_{\tau_2 \in \boldsymbol{\Lambda}_2} T\hat{\boldsymbol{\beta}}^b(\eta_2)' \boldsymbol{R}_1' [\boldsymbol{R}_1 \hat{\boldsymbol{V}}^b(\eta_2) \boldsymbol{R}_1']^{-1} \boldsymbol{R}_1 \hat{\boldsymbol{\beta}}^b(\eta_2), \tag{45}$$

where $\hat{\boldsymbol{\beta}}^b(\eta_j) = \left(\hat{\boldsymbol{\beta}}^{b\prime}_{(1)}(\eta_j), \hat{\boldsymbol{\beta}}^{b\prime}_{(2)}(\eta_j)\right)'$ are the bootstrap 2SLS counterparts of $\hat{\boldsymbol{\beta}}(\eta_j) = \left(\hat{\boldsymbol{\beta}}'_{(1)}(\eta_j), \hat{\boldsymbol{\beta}}'_{(2)}(\eta_j)\right)'$ for sub-sample $I_i(\eta_j)$. As for the original sample, the second break can be located before the first break ($j = 1$) or after the first break ($j = 2$). Let $\hat{\boldsymbol{Q}}^b_{(i)}(\eta_j) = T^{-1} \sum_{t \in I_i(\eta_j)} \hat{\boldsymbol{w}}^b_t \hat{\boldsymbol{w}}^{b\prime}_t$, $\hat{\boldsymbol{V}}^b(\eta_i) = \text{diag}\left(\hat{\boldsymbol{V}}^b_{(1)}(\eta_i), \hat{\boldsymbol{V}}^b_{(2)}(\eta_i)\right)$, $i, j = 1, 2$,

$$\hat{\boldsymbol{V}}^b_{(i)}(\eta_j) \;=\; \left(\hat{\boldsymbol{Q}}^b_{(i)}(\eta_j)\right)^{-1} \; \hat{\boldsymbol{M}}^b_{(i)}(\eta_j) \; \left(\hat{\boldsymbol{Q}}^b_{(i)}(\eta_j)\right)^{-1}, \tag{46}$$

and $\hat{\boldsymbol{M}}^b_{(i)}(\eta_j) = T^{-1} \sum_{t \in I_i(\eta_j)} \hat{\boldsymbol{\Upsilon}}^{b\prime}(t,T) \boldsymbol{z}^b_t \left(u^b_t + \boldsymbol{v}^{b\prime}_t \hat{\boldsymbol{\beta}}^b_{\boldsymbol{x},(i)}(\eta_j)\right) \left(\hat{\boldsymbol{\beta}}^{b\prime}_{\boldsymbol{x},(i)}(\eta_j) \boldsymbol{v}^b_t + u^b_t\right) \boldsymbol{z}^{b\prime}_t \hat{\boldsymbol{\Upsilon}}^b(t,T)$, $i, j = 1, 2$.

To show the validity of the WF bootstrap based on (43) we need the following assumption.

Assumption WR(b). *Assumption WR(a) holds with (vi) replaced by*

(vi) y^b_t *and* $\boldsymbol{x}^b_t$ *are obtained as in (36) and (37) with the bootstrap residuals such that* $u^b_{t,(i)} = \hat{u}_{t,(i)} \varsigma_{t,(i)}$ *and* $\boldsymbol{v}^b_{t,(i)} = \hat{\boldsymbol{v}}_{t,(i)} \varsigma_{t,(i)}$ *where* $\varsigma_{t,(i)} \overset{IID}{\sim} (0,1)$, $i = 1, 2$; $\mathrm{E}^b |\varsigma_t|^{4+\xi} = \bar{c} < \infty$, *for some* $\xi > 0$, *for all* t.

Assumption WR(b)(vi) states that the bootstrap residuals are obtained by multiplying the actual residuals by a random variable (independent of the data, with mean 0 and variance 1) in each of the two subsamples of residuals, i.e. $\{\hat{u}_{t,(1)}, \hat{\boldsymbol{v}}_{t,(1)}\}_{t=1}^{\hat{T}_1}$ and $\{\hat{u}_{t,(2)}, \hat{\boldsymbol{v}}_{t,(2)}\}_{t=\hat{T}_1+1}^{T}$.

5.1.3 IID fixed-design bootstrap (IF bootstrap)

(a) The null hypothesis $H_0 : m = 0$ against the alternative hypothesis $H_1 : m = k$.

Given the parameter estimates from RF, $\hat{\boldsymbol{\Delta}}(t,T)$, with $\hat{h} \geq 0$ estimated breaks, and the parameter estimates from SE, $\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1}$, under the null hypothesis $m = 0$, an IF bootstrap sample is obtained by adding to the estimated conditional mean where all the regressors are kept fixed (including the lagged variables) the bootstrap residuals obtained by resampling from the joint empirical distribution function of the centered 2SLS residuals:

$$\boldsymbol{x}^{b\prime}_t = \boldsymbol{z}'_t \hat{\boldsymbol{\Delta}}(t,T) + \boldsymbol{v}^{b\prime}_t, \quad y^b_t = \boldsymbol{x}^{b\prime}_t \hat{\boldsymbol{\beta}}_{\boldsymbol{x}} + \boldsymbol{z}'_{1,t} \hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1} + u^b_t = \boldsymbol{w}^{b\prime}_t \hat{\boldsymbol{\beta}} + u^b_t, \tag{47}$$

where $\boldsymbol{v}^{b\prime}_t$ and u^b_t are drawn as for the IR bootstrap case (a), $\boldsymbol{w}^b_t = \left(\boldsymbol{x}^{b\prime}_t, \boldsymbol{z}'_{1,t}\right)'$ which is different from $\boldsymbol{w}^b_t$ in (33) since $\boldsymbol{z}_{1,t}$ is kept fixed in the bootstrap. The regressors $\boldsymbol{z}_{1t}$ and $\boldsymbol{z}_t$ are as in (2) and (4). The IF bootstrap analogue of (10) is computed as in (34) but the bootstrap samples are generated as in (47), $\hat{\boldsymbol{w}}^b_t = \left(\hat{\boldsymbol{x}}^{b\prime}_t, \boldsymbol{z}'_{1,t}\right)'$, $\hat{\boldsymbol{x}}^{b\prime}_t = \boldsymbol{z}'_t \hat{\boldsymbol{\Delta}}^b(t,T)$, and $\hat{\boldsymbol{\Delta}}^b(t,T)$ is the bootstrap counterpart of $\hat{\boldsymbol{\Delta}}(t,T)$ with fixed regressors $\boldsymbol{z}_t$ and using the RF break-points $\hat{T}^*_j$ estimated in the original sample.

To show the asymptotic validity of the IF bootstrap based on sup-F test calculated as in (34), but with the bootstrap samples generated as in (47), we need the following assumption.

Assumption IF(a). *Assumption IR(a)(i)-(ii) holds and*

(iii) y^b_t *and* $\boldsymbol{x}^b_t$ *are obtained as in (47) with the bootstrap residuals such that* $u^b_t = \hat{u}_{\varsigma_t} - \bar{u}$, $\boldsymbol{v}^b_t = \hat{\boldsymbol{v}}_{\varsigma_t} - \bar{\boldsymbol{v}}$ *with* $\varsigma_t \overset{IID}{\sim} U(1,T)$, $t = 1, \ldots, T$.

Notice that Assumption IR(a)(iii) and Assumption IF(a)(iii) are identical, except that for the IF bootstrap $\boldsymbol{z}_t$ and $\boldsymbol{z}_{1,t}$ are kept fixed.

(b) The null hypothesis $H_0: m = 1$ against the alternative hypothesis $H_1: m = 2$.

Given the estimates of the RF estimated breaks $\hat{h} \geq 0$, the RF parameter estimates $\hat{\boldsymbol{\Delta}}(t,T)$, the SE break at $\hat{T}_1$ and the SE parameter estimates before the break, $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(1)}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(1)}$, and after the break, $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(2)}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(2)}$, the bootstrap samples are generated by adding to the estimated conditional mean (where all the regressors, including lagged regressors, are kept fixed) the bootstrap residuals obtained by resampling from the empirical distribution function of the centred 2SLS residuals, i.e.

$$\boldsymbol{x}_t^{b'} = \boldsymbol{z}_t'\hat{\boldsymbol{\Delta}}(t,T) + \boldsymbol{v}_t^{b'}, \tag{48}$$

$$y_t^b = 1_{t\leq\hat{T}_1}\left(\boldsymbol{x}_t^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(1)} + \boldsymbol{z}_{1,t}'\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(1)} + u_{t,(1)}^b\right) + 1_{t>\hat{T}_1}\left(\boldsymbol{x}_t^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(2)} + \boldsymbol{z}_{1,t}'\hat{\boldsymbol{\beta}}_{z_1,(2)} + u_{t,(2)}^b\right) = \boldsymbol{w}_t^{b\prime}\hat{\boldsymbol{\beta}}(t,T) + u_t^b, \tag{49}$$

where $u_{t,(i)}^b$ and $\boldsymbol{v}_{t,(i)}^b$ are drawn as for the IR bootstrap case (b). The IF bootstrap analogue of (16) is computed as in (40) with $\hat{\boldsymbol{w}}_t^b = (\hat{\boldsymbol{x}}_t^{b\prime}, \boldsymbol{z}_{1,t}')'$.

To show the asymptotic validity of the IF bootstrap of the sup-F test based on (40), but with the bootstrap samples generated as in (48)-(49), we need the following assumption.

Assumption IF(b). *Assumption IF(a) holds, except for (iii) which is replaced by*

(iii) y_t^b *and* $\boldsymbol{x}_t^b$ *are obtained as in (48) and (49) with the bootstrap residuals such that* $u_{t,(i)}^b = \hat{u}_{\varsigma_t,(i)} - \bar{u}_{(i)}$, $\boldsymbol{v}_{t,(i)}^b = \hat{\boldsymbol{v}}_{\varsigma_t,(i)} - \bar{\boldsymbol{v}}_{(i)}$, $i = 1,2$, *where* $\varsigma_{t,(1)} \overset{IID}{\sim} U(1,\hat{T}_1)$, $\varsigma_{t,(2)} \overset{IID}{\sim} U(1,T-\hat{T}_1)$.

Note again that Assumption IR(b)(iii) and Assumption IF(b)(iii) are identical, except that for the IF bootstrap $\boldsymbol{z}_t$ and $\boldsymbol{z}_{1,t}$ are kept fixed.

5.1.4 Wild fixed-design bootstrap (WF bootstrap)

(a) The null hypothesis $H_0: m = 0$ against the alternative hypothesis $H_1: m = k$.

The bootstrap samples are generated as in (47), with the regressors as in (2) and (4) respectively, and $\boldsymbol{v}_t^b$ and u_t^b are generated as for the WR bootstrap case (a). The WF bootstrap analogue of the statistic from (11) is calculated as in (41), with $\hat{\boldsymbol{Q}}_{(i)}^b = T^{-1}\sum_{t=T_{i-1}+1}^{T_i}\boldsymbol{w}_t^b\boldsymbol{w}_t^{b\prime}$ in (42), where $\boldsymbol{w}_t^b = (\hat{\boldsymbol{x}}_t^{b\prime}, \boldsymbol{z}_{1,t}')'$, and $\hat{\boldsymbol{M}}_{(i)}^b = T^{-1}\sum_{t\in I_i}\hat{\boldsymbol{\Upsilon}}^{b'}(t,T)\boldsymbol{z}_t\left(u_t^b + \boldsymbol{v}_t^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^b\right)\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^{b'}\boldsymbol{v}_t^b + u_t^b\right)\boldsymbol{z}_t'\hat{\boldsymbol{\Upsilon}}^b(t,T)$, the bootstrap analogue of $\hat{\boldsymbol{M}}_{(i)}$ from (14), $i = 1,\ldots,k+1$.

The WF bootstrap is valid under a less stringent assumption than Assumption WR(a). While for the WR bootstrap, we need to impose the equalities $\mathrm{E}^b((\boldsymbol{\varepsilon}_t^b\boldsymbol{\varepsilon}_t^{b'})\otimes(\boldsymbol{\varepsilon}_{t-i}^b\boldsymbol{\varepsilon}_{t-j}^{b'})) = 0$ and $\mathrm{E}^b((\boldsymbol{\varepsilon}_t^b\boldsymbol{\varepsilon}_t^{b'})\otimes(\boldsymbol{\varepsilon}_{t-i}^b\boldsymbol{r}_{t-j}')) = 0$ also for the original sample, for the WF bootstrap, due to its nonrecursive nature, we only need the values $\mathrm{E}^b((\boldsymbol{\varepsilon}_t^b\boldsymbol{\varepsilon}_t^{b'})\otimes(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-j}'))$ and $\mathrm{E}^b((\boldsymbol{\varepsilon}_t^b\boldsymbol{\varepsilon}_t^{b'})\otimes(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{r}_{t-j}'))$ to be mirrored in the original sample. Since both these latter values can be non-zero, we do not need to impose Assumption WR(a)(i) and (iii).

Assumption WF(a). *Assumptions 1, 4-11, 13(ii), 14 and Assumption WR(a)(v) hold. In addition,*

(i) $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{[Ts]}\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' \mid \mathcal{F}_t)\otimes(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-j}') = s\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon},ij}$ *in probability, uniformly in* s, *for all* t, $i \geq 1$, $j \geq 1$.

(ii) $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{[Ts]}\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' \mid \mathcal{F}_t)\otimes(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{r}_{t-j}') = s\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}\boldsymbol{r},ij}$ *in probability, uniformly in* s, *for all* t, $i \geq 1$, $j \geq 0$.

(iii) y_t^b *and* $\boldsymbol{x}_t^b$ *are obtained as in (47) with the bootstrap residuals such that* $u_t^b = \hat{u}_t\varsigma_t$ *and* $\boldsymbol{v}_t^b = \hat{\boldsymbol{v}}_t\varsigma_t$ *where* $\varsigma_t \overset{IID}{\sim}(0,1)$, $t = 1,\ldots,T$; $\mathrm{E}^b|\varsigma_t|^{4+\xi} = \bar{c} < \infty$, *for some* $\xi > 0$. *for all* t.

The WF bootstrap is more generally applicable than the WR bootstrap since Assumption WF(a)(i)-(ii) does not restrict Assumption 13(ii) by Assumption WR(a)(i), (iii), thus allowing for leverage effects in the form of an asymmetric response of volatility to positive and negative shocks of the same absolute magnitude, i.e. the popular EGARCH model. Assumption WF(a)(iii) and Assumption WR(a)(vi) are identical, except that for the WF bootstrap $\boldsymbol{z}_t$ and $\boldsymbol{z}_{1,t}$ are kept fixed.

(b) The null hypothesis $H_0 : m = 1$ against the alternative hypothesis $H_1 : m = 2$.

The bootstrap samples are generated as for the IF bootstrap case (b) with the bootstrap residuals drawn as for the WR bootstrap case (b). The WF bootstrap analogue of (18) is calculated as in (43) with $\boldsymbol{w}_t^b = (\hat{\boldsymbol{x}}_t^{b\prime}, \boldsymbol{z}_{1,t}')'$ (which is different from $\hat{\boldsymbol{w}}_t$ in case (a) since it takes into account the break at $\hat{T}_1$), and $\hat{\boldsymbol{M}}_{(i)}^b(\eta_j) = T^{-1}\sum_{t\in I_i(\eta_j)} \hat{\boldsymbol{\Upsilon}}^{b'}(t,T)\boldsymbol{z}_t \left(u_t^b + \boldsymbol{v}_t^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^b(\eta_j)\right)\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^{b\prime}(\eta_j)\boldsymbol{v}_t^b + u_t^b\right)\boldsymbol{z}_t'\hat{\boldsymbol{\Upsilon}}^b(t,T)$, $i,j = 1,2$.

To show the validity of the WF bootstrap based on (43), but with the bootstrap samples generated as mentioned above, we need the following assumption.

Assumption WF(b). *Assumption WF(a) holds, except that (iii) is replaced by*

(iii) y_t^b and $\boldsymbol{x}_t^b$ are obtained as in (49) and (48) with the bootstrap residuals such that $u_{t,(i)}^b = \hat{u}_{t,(i)}\varsigma_{t,(i)}$ and $\boldsymbol{v}_{t,(i)}^b = \hat{\boldsymbol{v}}_{t,(i)}\varsigma_{t,(i)}$ where $\varsigma_{t,(i)} \overset{IID}{\sim} (0,1)$, $i = 1,2$; $\mathrm{E}^b|\varsigma_t|^{4+\xi} < \infty$, for some $\xi > 0$. for all t.

Assumption WF(b)(iii) and Assumption WR(b)(vi) are identical, except that for the WF bootstrap $\boldsymbol{z}_t$ and $\boldsymbol{z}_{1,t}$ are kept fixed.

5.2 Validity of bootstrap tests

In the following theorems, we prove that the difference between the bootstrap distribution and the asymptotic distribution of the sup-F and sup-$Wald$ statistics converges uniformly in probability to zero.

Theorem 5. *Under Assumption IR(a) for the IR bootstrap or Assumption IF(a) for the IF bootstrap:*

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

as $T \to \infty$, where P^b denotes the probability measure induced by the IR bootstrap or the IF bootstrap, sup-F_T *is given in* (10)*, and* sup-F_T^b *is given in* (34) *and computed as described in Assumption IR(a)(iii) for the IR bootstrap or as described in Assumption IF(a)(iii) for the IF bootstrap.*

Theorem 5 shows that under Assumption IR(a) or Assumption IF(a) the statistics sup-F_T^b and sup-F_T have the same asymptotic distribution given in Theorem 1.

Theorem 6. *Under Assumption WR(a) for the WR bootstrap or Assumption WF(a) for the WF bootstrap:*

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

as $T \to \infty$, where P^b denotes the probability measure induced by the WR bootstrap or the WF bootstrap, sup-$Wald_T$ *is given in* (11)*, and* sup-$Wald_T^b$ *is given in* (41) *and computed as described in Assumption WR(a)(vi) for the WR bootstrap or as described in Assumption WF(a)(iii) for the WF bootstrap.*

Theorem 6 shows that under Assumption WR(a) or Assumption WF(a) the statistics sup-$Wald_T^b$ and sup-$Wald_T$ have the same asymptotic distribution given in Theorem 2.

Remark 1. *Under Assumption WF(a),* $\sup\text{-}Wald_T \Rightarrow \sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}_\epsilon}\left\{\tilde{\boldsymbol{K}}^{*\prime}(\boldsymbol{\lambda};\boldsymbol{\pi}^0)\ [\boldsymbol{H}^*(\boldsymbol{\lambda})\otimes\boldsymbol{V}^*]\ \tilde{\boldsymbol{K}}^*(\boldsymbol{\lambda};\boldsymbol{\pi}^0)\right\}$ *with* $\boldsymbol{V}^* = \boldsymbol{\Upsilon}^0(\boldsymbol{\Upsilon}^{0\prime}\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Upsilon}^{0\prime}$*, where* $\tilde{\boldsymbol{\Sigma}}_1$ *is as in Definition B2 of the Appendix B below, and the covariance of the Gaussian process* $\tilde{\boldsymbol{K}}^*(\boldsymbol{\lambda};\boldsymbol{\pi}^0)$ *satisfies Assumption 13(ii) (included in Assumption WF(a)), namely* $\tilde{\boldsymbol{K}}^*(\boldsymbol{\lambda};\boldsymbol{\pi}^0)$ *is defined as in Definition 3 with* $\tilde{\boldsymbol{D}}(\cdot)\equiv\tilde{\boldsymbol{N}}_1'\boldsymbol{W}(\cdot)$*,* $\tilde{\boldsymbol{D}}^*(\cdot)\equiv\tilde{\tilde{\boldsymbol{W}}}(\cdot)\boldsymbol{\beta}_{\boldsymbol{x}}^0$ *where* $\tilde{\boldsymbol{N}}_1'\tilde{\boldsymbol{N}}_1=\tilde{\boldsymbol{\Sigma}}_u$*,* $\tilde{\boldsymbol{N}}_1'\tilde{\boldsymbol{N}}_2=\tilde{\boldsymbol{\Sigma}}_{u\boldsymbol{v}}$ *and* $\tilde{\boldsymbol{N}}_2'\tilde{\boldsymbol{N}}_2=\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}}$*; see the definitions just after (28) and replace* $\boldsymbol{\Sigma}_u$*,* $\boldsymbol{\Sigma}_{u\boldsymbol{v}}$ *and* $\boldsymbol{\Sigma}_{\boldsymbol{v}}$ *with* $\tilde{\boldsymbol{\Sigma}}_u$*,* $\tilde{\boldsymbol{\Sigma}}_{u\boldsymbol{v}}$ *and* $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}}$*.*

Remark 2. *Under Assumption WR(a),* $\sup\text{-}Wald_T \Rightarrow \sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}_\epsilon}\left\{\tilde{\boldsymbol{K}}^{*\prime}(\boldsymbol{\lambda};\boldsymbol{\pi}^0)\ [\boldsymbol{H}^*(\boldsymbol{\lambda})\otimes\boldsymbol{V}^*]\ \tilde{\boldsymbol{K}}^*(\boldsymbol{\lambda};\boldsymbol{\pi}^0)\right\}$ *with* $\boldsymbol{V}^* = \boldsymbol{\Upsilon}^0(\boldsymbol{\Upsilon}^{0\prime}\tilde{\tilde{\boldsymbol{\Sigma}}}_1\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Upsilon}^{0\prime}$ *where* $\tilde{\tilde{\boldsymbol{\Sigma}}}_1$ *is as in Definition B2 of the Appendix B, and the covariance of the Gaussian process* $\tilde{\boldsymbol{K}}^*(\boldsymbol{\lambda};\boldsymbol{\pi}^0)$ *satisfies Assumption WR(a), namely* $\tilde{\boldsymbol{K}}^*(\boldsymbol{\lambda};\boldsymbol{\pi}^0)$ *is defined as in Definition 3 with* $\tilde{\boldsymbol{D}}(\cdot)\equiv\tilde{\boldsymbol{N}}_1'\boldsymbol{W}(\cdot)$*,* $\tilde{\boldsymbol{D}}^*(\cdot)\equiv\tilde{\tilde{\boldsymbol{W}}}(\cdot)\boldsymbol{\beta}_{\boldsymbol{x}}^0$ *where* $\tilde{\boldsymbol{N}}_1'\tilde{\boldsymbol{N}}_1=\tilde{\tilde{\boldsymbol{\Sigma}}}_u$*,* $\tilde{\boldsymbol{N}}_1'\tilde{\boldsymbol{N}}_2=\tilde{\tilde{\boldsymbol{\Sigma}}}_{u\boldsymbol{v}}$ *and* $\tilde{\boldsymbol{N}}_2'\tilde{\boldsymbol{N}}_2=\tilde{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{v}}$*; see the definitions just after (28) and replace* $\boldsymbol{\Sigma}_u$*,* $\boldsymbol{\Sigma}_{u\boldsymbol{v}}$ *and* $\boldsymbol{\Sigma}_{\boldsymbol{v}}$ *with* $\tilde{\tilde{\boldsymbol{\Sigma}}}_u$*,* $\tilde{\tilde{\boldsymbol{\Sigma}}}_{u\boldsymbol{v}}$ *and* $\tilde{\tilde{\boldsymbol{\Sigma}}}$*.*

Theorem 7. *Under Assumption IR(b) for the IR bootstrap and Assumption IF(b) for the IF bootstrap:*

$$\sup_{c\in\mathbb{R}}\left|P^b\left(\sup\text{-}F_T^b(2|1)\le c\right)-P(\sup\text{-}F_T(2|1)\le c)\right|\xrightarrow{p}0$$

as $T\to\infty$*, where* P^b *denotes the probability measure induced by the IR bootstrap or the IF bootstrap,* $F_T(2|1)$ *is given in (16), and* $\sup\text{-}F_T^b(2|1)$ *is given in (40) and computed as described in Assumption IR(b)(iii) for the IR bootstrap or as described in Assumption IF(b)(iii) for the IF bootstrap.*

Theorem 7 shows that under Assumption IR(b) or Assumption IF(b) the statistics $F_T(2|1)$ and $F_T^b(2|1)$ have the same asymptotic distribution given in Theorem 3.

Theorem 8. *Under Assumption WR(b) for the WR bootstrap and Assumption WF(b) for the WF bootstrap:*

$$\sup_{c\in\mathbb{R}}\left|P^b\left(\sup\text{-}Wald_T^b(2|1)\le c\right)-P(\sup\text{-}Wald_T(2|1)\le c)\right|\to 0$$

as $T\to\infty$*, where* P^b *denotes the probability measure induced by the WR bootstrap or the WF bootstrap,* $\sup\text{-}Wald_T(2|1)$ *is given in (18),* $\sup\text{-}Wald_T^b(2|1)$ *is given in (43) which is computed as described in Assumption WR(b)(vi) for the WR bootstrap or as described in Assumption WF(b)(iii) for the WF bootstrap.*

Remark 3. *Theorem 8 shows that under Assumption WR(b) or Assumption WF(b) the statistics* $Wald_T(2|1)$ *and* $Wald_T^b(2|1)$ *have the same asymptotic distribution given in Theorem 4. For Theorem 8, the Gaussian process* $\tilde{\boldsymbol{K}}_i^*(\eta_i;\boldsymbol{\pi}^0,\lambda_1^0)$*,* $i=1,2$*, from Theorem 4, is redefined such that* $\tilde{\boldsymbol{D}}(\cdot)\equiv\tilde{\boldsymbol{N}}_1'\boldsymbol{W}(\cdot)$*,* $\tilde{\boldsymbol{D}}^*(\cdot)\equiv\tilde{\tilde{\boldsymbol{W}}}(\cdot)\boldsymbol{\beta}_{\boldsymbol{x}}^0$ *with (i)* $\tilde{\boldsymbol{N}}_1'\tilde{\boldsymbol{N}}_1=\tilde{\boldsymbol{\Sigma}}_u$*,* $\tilde{\boldsymbol{N}}_1'\tilde{\boldsymbol{N}}_2=\tilde{\boldsymbol{\Sigma}}_{u\boldsymbol{v}}$ *and* $\tilde{\boldsymbol{N}}_2'\tilde{\boldsymbol{N}}_2=\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}}$ *under Assumption 13(ii) included in Assumption WF(b), see the definitions just after (28) and replace* $\boldsymbol{\Sigma}_u$*,* $\boldsymbol{\Sigma}_{u\boldsymbol{v}}$ *and* $\boldsymbol{\Sigma}_{\boldsymbol{v}}$ *with* $\tilde{\boldsymbol{\Sigma}}_u$*,* $\tilde{\boldsymbol{\Sigma}}_{u\boldsymbol{v}}$ *and* $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}}$*; and with (ii)* $\tilde{\boldsymbol{N}}_1'\tilde{\boldsymbol{N}}_1=\tilde{\tilde{\boldsymbol{\Sigma}}}_u$*,* $\tilde{\boldsymbol{N}}_1'\tilde{\boldsymbol{N}}_2=\tilde{\tilde{\boldsymbol{\Sigma}}}_{u\boldsymbol{v}}$ *and* $\tilde{\boldsymbol{N}}_2'\tilde{\boldsymbol{N}}_2=\tilde{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{v}}$ *under Assumption WR(a)(i)-(iii) which is part of Assumption WR(b) and it restricts Assumption 13(ii), see the definitions just after (28) and replace* $\boldsymbol{\Sigma}_u$*,* $\boldsymbol{\Sigma}_{u\boldsymbol{v}}$ *and* $\boldsymbol{\Sigma}_{\boldsymbol{v}}$ *with* $\tilde{\tilde{\boldsymbol{\Sigma}}}_u$*,* $\tilde{\tilde{\boldsymbol{\Sigma}}}_{u\boldsymbol{v}}$ *and* $\tilde{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{v}}$*.*

Remark 4. *The proofs of Theorems 5-8 also include the proofs for the validity of the bootstraps for parameter estimates under the null hypotheses considered. See for example the arguments in the Supplemental Appendix leading to (2.97), (2.98), (2.109) from the proof of Theorem 5 or the arguments leading to (2.135)-(2.136) from the proof of Theorem 7.*

Remark 5. *When there are no endogenous regressors in SE, then Theorems 1-4 hold with* $\boldsymbol{w}_t=\boldsymbol{z}_{1,t}$ *instead of* $\boldsymbol{w}_t=(\boldsymbol{x}_t',\boldsymbol{z}_{1,t}')'$*. In particular, the asymptotic distributions of the* $\sup\text{-}F_T$*,* $\sup\text{-}F_T(2|1)$*,* $\sup\text{-}Wald_T$ *and*

sup-$Wald_T(2|1)$ *from Theorems 1-4 are as in BP and are generated exclusively by partial sums of* $\{\boldsymbol{z}_{1,t}u_t\}$. *The IR and WR bootstrap samples are obtained by recursively generating* y_t^b *(only):* $y_t^b = \boldsymbol{z}_{1,t}^{b\prime}\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1} + u_t^b$ *for testing* $H_0 : m = 0$ *against* $H_1 : m = k$ *with* $k = 1, 2$, *or* $y_t^b = 1_{t\leq\hat{T}_1}\left(\boldsymbol{z}_{1,t}^{b\prime}\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(1)} + u_{t,(1)}^b\right) + 1_{t>\hat{T}_1}\left(\boldsymbol{z}_{1,t}^{b\prime}\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(2)} + u_{t,(2)}^b\right)$ *for testing* $H_0 : m = 1$ *against* $H_1 : m = 2$, *where* $\boldsymbol{z}_{1,t}^b = \left(\boldsymbol{r}_{1,t}', y_{t-1}^b, \boldsymbol{x}_{t-1}', y_{t-2}^b, \boldsymbol{x}_{t-2}', \ldots, \boldsymbol{x}_{t-\tilde{q}_1}'\right)'$. *The IF and WF bootstrap samples are obtained by only adding to the conditional mean of* y_t *the IF and WF bootstrap residuals, i.e.* $y_t^b = \boldsymbol{z}_{1,t}'\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1} + u_t^b$ *for testing* $H_0 : m = 0$ *against* $H_1 : m = k$ *with* $k = 1, 2$, *or* $y_t^b = 1_{t\leq\hat{T}_1}\left(\boldsymbol{z}_{1,t}'\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(1)} + u_{t,(1)}^b\right) + 1_{t>\hat{T}_1}\left(\boldsymbol{z}_{1,t}\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(2)} + u_{t,(2)}^b\right)$ *for testing* $H_0 : m = 1$ *against* $H_1 : m = 2$, *where* $\boldsymbol{z}_{1,t} = \left(\boldsymbol{r}_{1,t}', y_{t-1}, \boldsymbol{x}_{t-1}', y_{t-2}, \boldsymbol{x}_{t-2}', \ldots, \boldsymbol{x}_{t-\tilde{q}_1}'\right)'$, *and for all bootstraps* $\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1}$, $\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(1)}$, $\hat{\boldsymbol{\beta}}_{\boldsymbol{z}_1,(2)}$ *are the parameter estimates under the null hypothesis. With these changes, the Theorems 5-8 hold with* $\boldsymbol{w}_t = \boldsymbol{z}_{1,t}$ *instead of* $\boldsymbol{w}_t = (\boldsymbol{x}_t', \boldsymbol{z}_{1,t}')'$.

Remark 6. *When there are endogenous regressors in SE, but the RF is stable, then Theorems 1-4 hold with* $\hat{\boldsymbol{x}}_t$ *computed over the full sample, i.e.* $\hat{\boldsymbol{x}}_t = \boldsymbol{z}_t'\hat{\boldsymbol{\Delta}}$ *and* $\hat{\boldsymbol{\Delta}} = (\sum_{t=1}^T \boldsymbol{z}_t\boldsymbol{z}_t')^{-1}\sum_{t=1}^T \boldsymbol{z}_t\boldsymbol{x}_t'$. *Also, Theorems 5-8 hold with* $\boldsymbol{x}_t^{b\prime} = \boldsymbol{z}_t^{b\prime}\hat{\boldsymbol{\Delta}} + \boldsymbol{v}_t^{b\prime}$ *for the IR and WR bootstraps, and with* $\boldsymbol{x}_t^{b\prime} = \boldsymbol{z}_t'\hat{\boldsymbol{\Delta}} + \boldsymbol{v}_t^{b\prime}$ *for the IF and WF bootstraps.*

Remark 7. *Theorems 1-8 can be straightforwardly generalised to more than two SE breaks.*

6 Simulation study

In this section, we investigate the finite sample performance of the bootstrap versions of the sup-F and sup-$Wald$ statistics described in Section 5. We consider a number of designs that involve stability or instability in the SE and/or the RF. In each case, the performance of the bootstrap tests is compared to that of the corresponding tests based on the limiting distributions tabulated by BP. It should be noted that the latter limiting distributions are only valid in scenarios involving stable RFs. For the designs with unstable RFs, we also report rejection frequencies for asymptotic tests based on the correct limiting distribution derived in Section 4.[24]

Specifically, we consider the following scenarios:

1) There is no break in RF and no break in the structural equation SE. We test $H_0 : 0$ break in SE against $H_1 : 1$ break in SE. The data generating process (DGP) is as follows:

$$x_t = \alpha_x + \boldsymbol{r}_t'\boldsymbol{\delta}_r^0 + \delta_{x_1}^0 x_{t-1} + \delta_{y_1}^0 y_{t-1} + v_t, \quad \text{for } t = 1, \ldots, T, \tag{50}$$

$$y_t = \alpha_y + x_t\beta_x^0 + \beta_{r_1}^0 r_{1,t} + \beta_{y_1}^0 y_{t-1} + u_t, \quad \text{for } t = 1, \ldots, T. \tag{51}$$

We consider $\alpha_x = \alpha_y = 1$, $\boldsymbol{\delta}_r^0 = (1.5, 1.5, 1.5, 1.5)$ is a 4×1 parameter vector; $\boldsymbol{r}_t = (r_{1,t}, \boldsymbol{r}_{2,t}')'$, $\boldsymbol{r}_t \overset{IID}{\sim} N(\boldsymbol{0}_4, \boldsymbol{I}_4)$; $\beta_x^0 = -0.6$, $\beta_{r_1}^0 = 1.5$; $\delta_{x_1}^0 = \delta_{y_1}^0 = 0.1$, $\beta_{y_1}^0 = -0.8$. For these parameter values, the VARX(1,0) corresponding to (50)-(51) has roots outside the unit circle: 1.17 and -10.67. We assume that the error terms $\boldsymbol{\varepsilon}_t = (u_t, v_t)'$ are:

(A) homoskedastic symmetric: u_t and $v_t \overset{IID}{\sim} N(0,1)$, $\text{cov}(u_t, v_t) = 0.5$, $t = 1, \ldots, T$.

(B) homoskedastic skewed: $u_t = (\varepsilon_t^u - 2)/2$, $v_t = (\varepsilon_t^v - 2)/2$, ε_t^u and $\varepsilon_t^v \overset{IID}{\sim} \chi_2(2)$, $\text{cov}(u_t, v_t) = 0.5$, $t = 1, \ldots, T$.

(C) conditional heteroskedastic: $\tilde{u}_t = \sigma_{\tilde{u},t}\xi_{\tilde{u},t}$ and $\tilde{v}_t = \sigma_{\tilde{v},t}\xi_{\tilde{v},t}$ are GARCH(1,1) processes with $\xi_{\tilde{u},t}$ and $\xi_{\tilde{v},t} \overset{IID}{\sim} N(0,1)$, $\text{cov}(\xi_{\tilde{u},t}, \xi_{\tilde{v},t}) = 0.5$, $\sigma_{\tilde{u},t}^2 = \gamma_0 + \gamma_1\tilde{u}_{t-1}^2 + \gamma_2\sigma_{\tilde{u},t-1}^2$, $\sigma_{\tilde{v},t}^2 = \gamma_0 + \gamma_1\tilde{v}_{t-1}^2 + \gamma_2\sigma_{\tilde{v},t-1}^2$,

[24] The critical values are obtained by simulating the sampling distribution of the test statistic in question in samples of size 1000 and are based on 10,000 replications.

where $\gamma_0 = \gamma_1 = \gamma_2 = 0.1$. We standardize $e_t = (u_t, v_t)$ by the unconditional variance $\text{var}(u_t) = \text{var}(v_t) = \gamma_0/(1-\gamma_1-\gamma_2)$. Hence we consider $u_t = \tilde{u}_t/\sqrt{\text{var}(u_t)}$ and $v_t = \tilde{v}_t/\sqrt{\text{var}(v_t)}$, $T = 1, \ldots, T$.

For cases (A) and (B) we test H_0 : 0 break in SE against H_1 : 1 break in SE using the sup-F test based on the critical values of the asymptotic distribution derived in BP and Bai and Perron (2003)[25], and the critical values obtained using the IR and IF bootstraps.

For the cases (A) and (C) we test H_0 : 0 break in SE against H_1 : 1 break in SE using the sup-$Wald$ test based on the critical values in BP,[26] and those obtained by the WR and WF bootstraps.

2) There is a break in RF at $[T/4]$ and no break in the structural equation (SE). We test H_0 : 0 break in SE against H_1 : 1 break in SE. The DGP is as follows:

$$\begin{aligned} x_t &= \alpha_x + \boldsymbol{r}_t'\boldsymbol{\delta}_r^0 + \delta_{x_1}^0 x_{t-1} + \delta_{y_1}^0 y_{t-1} + v_t, \qquad \text{for } t = 1, \ldots, [T/4], && (52)\\ &= -\alpha_x - \boldsymbol{r}_t'\boldsymbol{\delta}_r^0 - \delta_{x_1}^0 x_{t-1} - \delta_{y_1}^0 y_{t-1} + v_t, \qquad \text{for } t = [T/4]+1, \ldots, T, && (53)\\ y_t &= \alpha_y + x_t\beta_x^0 + \beta_{r_1}^0 r_{1,t} + \beta_{y_1}^0 y_{t-1} + u_t, \qquad \text{for } t = 1, \ldots, T. && (54) \end{aligned}$$

We consider the same parameter values as in scenario 1). For the sup-F test we consider the errors as in cases (A) and (B), and compute the rejection frequencies based on the BP critical values (which do not take into account the break in RF), BCH critical values (based on the approximation of the asymptotic distribution from Theorem 1), and the IR and IF bootstraps. For the sup-$Wald$ test we consider the errors as in cases (A) and (C), and compute the rejection frequencies based on the BP critical values (which again do not take into account the break in RF), BCH critical values (based on the approximation of the asymptotic distribution from Theorem 2), and the WR and WF bootstraps.

3) There is no break in RF, but there is a break in SE at $[3T/4]$. We test H_0 : 1 break in SE against H_1 : 2 breaks in SE. The DGP is as follows:

$$\begin{aligned} x_t &= \alpha_x + \boldsymbol{r}_t'\boldsymbol{\delta}_r^0 + \delta_{x_1}^0 x_{t-1} + \delta_{y_1}^0 y_{t-1} + v_t, \quad \text{for } t = 1, \ldots, T, && (55)\\ y_t &= \alpha_y - x_t\beta_x^0 + \beta_{r_1}^0 r_{1,t} - \beta_{y_1}^0 y_{t-1} + u_t, \qquad \text{for } t = 1, \ldots, [3T/4], && (56)\\ &= -\alpha_y + x_t\beta_x^0 - \beta_{r_1}^0 r_{1,t} + \beta_{y_1}^0 y_{t-1} + u_t, \qquad \text{for } t = [3T/4]+1, \ldots, T. && (57) \end{aligned}$$

We consider the same parameter values as in scenario 1). For the sup-F test the errors are as in cases (A) and (B) for which we compute the rejection frequencies based on the BP critical values, and the IR and IF bootstraps. For the sup-$Wald$ test, the errors are as in cases (A) and (C) for which we compute rejection frequencies based on the BP critical values, and the WR and WF bootstraps.

4) There is a break in RF at $[T/4]$ and a break in SE at $[3T/4]$. We test H_0 : 1 break in SE against H_1 : 2 breaks in SE. The DGP is as follows:

$$\begin{aligned} x_t &= \alpha_x + \boldsymbol{r}_t'\boldsymbol{\delta}_r^0 + \delta_{x_1}^0 x_{t-1} + \delta_{y_1}^0 y_{t-1} + v_t, \qquad \text{for } t = 1, \ldots, [T/4], && (58)\\ &= -\alpha_x - \boldsymbol{r}_t'\boldsymbol{\delta}_r^0 - \delta_{x_1}^0 x_{t-1} - \delta_{y_1}^0 y_{t-1} + v_t, \qquad \text{for } t = [T/4]+1, \ldots, T, && (59)\\ y_t &= \alpha_y - x_t\beta_x^0 + \beta_{r_1}^0 r_{1,t} - \beta_{y_1}^0 y_{t-1} + u_t, \qquad \text{for } t = 1, \ldots, [3T/4], && (60)\\ &= -\alpha_y + x_t\beta_x^0 - \beta_{r_1}^0 r_{1,t} + \beta_{y_1}^0 y_{t-1} + u_t, \qquad \text{for } t = [3T/4]+1, \ldots, T. && (61) \end{aligned}$$

[25] We use their Table 2c with $q = 4$. Note that their nominal value is $\alpha = 1 - \alpha_1$, while $\alpha_1 = 0.01, 0.05, 0.1$ denote our nominal values.

[26] Note that the BP critical values for sup-$Wald$ are the same as the critical for the sup-F.

We consider the same parameter values as in scenario 1). For the sup-F test we compute the rejection frequencies based on the BP critical values (which do not take into account the break in RF), BCH critical values (based on the approximation of the asymptotic distribution from Theorem 3), and the IR and IF bootstraps. For the sup-$Wald$ test we compute the rejection frequencies based on the BP critical values (which do not take into account the break in RF), BCH critical values (based on the approximation of the asymptotic distribution from Theorem 4), and the WR and WF bootstraps.

For the WR and WF bootstraps the auxiliary distribution (from Assumptions WR(a), WR(b), WF(a), WF(b)) is the Rademacher distribution proposed by Liu (1988) which assigns 0.5 probability to the value $\varsigma_t = -1$ and 0.5 probability to $\varsigma_t = 1$, $t = 1, \ldots, T$. The same ς_t is used to obtain both the bootstrap residuals $u_t^b = \hat{u}_t \varsigma_t$ and the bootstrap residuals $v_t^b = \hat{v}_t \varsigma_t$ in order to preserve the contemporaneous correlation between the error terms. For the IR and IF bootstraps, the bootstrap residuals u_t^b and v_t^b are obtained by resampling with replacement from the joint distribution of the centered residuals $\left(\hat{u}_t - T^{-1}\sum_{t=1}^{T}\hat{u}_t\right)$ and $\left(\hat{v}_t - T^{-1}\sum_{t=1}^{T}\hat{v}_t\right)$. Each scenario 1)-4) was repeated $N = 10,000$ times considering $T = 120, 240, 480$ for the sample size and $B = 399$ bootstrap replications.

In Tables 1-8, we report the rejection rates of the IR, IF, WR and WF bootstraps for scenarios 1)-4): $N^{-1}\sum_{j=1}^{N} 1_{t_j \geq t^b_{1-\alpha_1,j}}$, where $\alpha_1 = 0.10, 0.05, 0.01$ are the nominal values of the tests; t_j is the statistic (sup-F or 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 $j = 1, \ldots, N$. The rejection rates of the asymptotic test in Tables 1-8 from Appendix C are computed similarly, with $t^b_{1-\alpha_1,j}$ replaced by the $1-\alpha_1$ quantile of the asymptotic distribution approximated using a large sample, $T = 1000$, $N = 10000$, the errors taken as in case (A), and assuming the date of the breaks in RF and SE is known.

For the sup-$Wald$ tests, the sandwich covariance matrix $\hat{V}(\boldsymbol{\lambda})$ (from (13) and (21)) and its bootstrap counterpart $\hat{V}^b(\boldsymbol{\lambda})$ are estimated using a heteroskedasticity-consistent covariance matrix estimator (HCCME); see e.g. Davidson and MacKinnon (1993) pp.552-554. We have tried all four versions of the HCCME in Davidson and MacKinnon (1993): HC_0, HC_1, HC_2, HC_3, all giving similar results. In the tables we report the results for HCCME-HC_0. We also considered HAC estimators of the sandwich matrix, but their results are very poor in finite sample and are omitted.

For the IR and WR bootstraps, the bootstrap samples were generated recursively with start-up values for y_1^b and x_1^b being given by the first observations from the sample (x_1, y_1); see Davidson and MacKinnon (1993). We have also considered the case when the start-up values are drawn at random from the full sample of y_t's and x_t's; see e.g. Berkowitz and Kilian (2000), but the results are similar.

In all tables, the bootstrap samples are generated by imposing the true null hypothesis. We also impose the true number of breaks in RF (zero breaks for scenarios 1) and 3), and one break which we estimate, for scenarios 2) and 4)). For scenarios 3) and 4), the break in SE is estimated (but not tested for) prior to testing for the second break in SE. We have experimented with the situation in which the first SE break is estimated and tested for (H_0: 0 SE break against H_1: 1 SE break) before testing H_0: 1 SE break against H_1 : 2 SE breaks, and the test had very good power (these results are not reported). The value of ϵ, the cut-off from Assumption 6, was taken equal to 0.15 which is a typical value used in the literature.

Regarding the sup-F test, as it can be seen from Tables 1-4, the IR bootstrap works the best both when the error term is homoskedastic and from a symmetric distribution (case (A)) and when it homoskedastic but from a skewed distribution (case (B)). The IF bootstrap performs satisfactorily, but less well than the IR bootstrap

since it does not take into account the recursive nature in the DGP.

In Tables 1 and 3, the RF is stable and so the BP critical values are valid asymptotically: for testing no breaks versus one, the asymptotic test based on BP critical values performs well (Table 1), but for the test of one versus two breaks, the asymptotic test only has empirical size close to the nominal level in the largest sample (Table 3). If the RF is unstable then the BP critical values are invalid, and their (incorrect) use in testing no breaks versus one leads to over-rejection (Table 2), and in testing for one versus two breaks leads to under-rejection (Table 4). When the RF is unstable, valid asymptotic tests can be performed using the critical values derived from the limiting distributions presented in Section 4. These asymptotic tests possess empirical size equal or very close to the nominal size in all sample sizes for the test of zero versus 1 break (Table 2), and in the largest sample size for the test of one versus two breaks (Table 4); in each case see the column headed 'BCH asymptotic'.

Regarding the sup-$Wald$ test, as it can be seen from Tables 5-8, the WR and WF bootstraps work very well as opposed to the asymptotic tests (BP and BCH) which have large rejection probabilities. In Tables 6 and 8 (when there is a break in RF), the BCH distribution gives better results than the BP distribution as expected. In particular, it can be noticed from these tables that the actual rejection probabilities based on the BCH distribution decrease faster than those based on the BP distribution as the sample size increases.

In summary, within the simulation designs we consider, our results make a compelling case for the use of the bootstrap-based tests.

7 Conclusion

In this paper, we prove the asymptotic validity of the (IID and wild) recursive bootstrap and of the (IID and wild) fixed-regressor bootstrap for the most popular structural changes tests in dynamic models estimated by ordinary least squares (OLS) or by two-stage least squares (2SLS) with stable or unstable first stages, when the location of the break is unknown. This important result has been missing from the literature, although these bootstraps have been extensively used in many applications in the past few decades. In addition, we derive the asymptotic distribution of the structural break tests when the model is estimated by 2SLS and the first-stage equation is unstable. We show that this distribution is different from the one derived in Bai and Perron (1998) based on OLS and from the one in Hall, Han, and Boldea (2012) based on 2SLS with stable first-stage equation, since it depends on the number of breaks in the first-stage equation and their relative position. Our simulation results indicate that these bootstraps perform better than the asymptotic tests of BP/HHB and of this paper, in particular in the presence of conditional heteroskedasticity.

A Appendix: VARX and VMAX representations of (1) and (3)

To derive the results in this paper, we use a VARX representation of the SE and RF. In this section we give the VARX representation of (1) and (3) for the **Case (I)** when there are no breaks in SE ($m = 0$), but there are breaks in RF ($h > 0$), and **Case (II)** when there are breaks in both SE and RF. For clarity, in **Case (II)** we focus on the scenario in which there is a break in SE ($m = 1$) and there are breaks in RF ($h > 0$), and the break in SE could coincide with one of the breaks in RF, but our results also extend to the case when $m > 1$.

Case(I). When there is no break in SE ($m = 0$), but there are breaks in RF ($h > 0$), we can write (1) and

(3) in the following VARX($\tilde{p}$,0) form:[27]

$$\boldsymbol{A}_0\tilde{\boldsymbol{y}}_t = \sum_{i=1}^{\tilde{p}} \boldsymbol{A}_i(t,T)\tilde{\boldsymbol{y}}_{t-i} + \mathbf{B}(t,T)\boldsymbol{r}_t + \boldsymbol{\varepsilon}_t, \tag{62}$$

with $\boldsymbol{A}_i(t,T) = \sum_{j=1}^{h+1} \boldsymbol{A}_{i,(j)} 1_{t\in I_j^*}$ and $\mathbf{B}(t,T) = \sum_{j=1}^{h+1} \mathbf{B}_{(j)} 1_{t\in I_j^*}$, where I_j^* is the j^{th} segment of the true RF h-partition $\boldsymbol{\pi}^0$, i.e. $I_j^* = (T_{j-1}^* + 1, \ldots, T_j^*) = ([T\pi_{j-1}^0] + 1, \ldots, [T\pi_j^0])$, $j = 1, \ldots, h+1$, and

$$\boldsymbol{A}_0 = \begin{pmatrix} 1 & -\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime} \\ \mathbf{0}_{p_1} & \boldsymbol{I}_{p_1} \end{pmatrix}, \ \boldsymbol{A}_{i,(j)} = \begin{pmatrix} \beta_{y_i}^0 & \boldsymbol{\beta}_{\boldsymbol{x}_i}^{0\prime} \\ \boldsymbol{\delta}_{y_i,(j)}^0 & \boldsymbol{\Delta}_{\boldsymbol{x}_i,(j)}^{0\prime} \end{pmatrix}, \ \mathbf{B}_{(j)} = \begin{pmatrix} \boldsymbol{\beta}_{\boldsymbol{r}_1}^{0\prime} & \mathbf{0}_{q_2}' \\ \boldsymbol{\Delta}_{\boldsymbol{r}_1,(j)}^{0\prime} & \boldsymbol{\Delta}_{\boldsymbol{r}_2,(j)}^{0\prime} \end{pmatrix}, \tag{63}$$

where the matrices of coefficients $\boldsymbol{A}_0$ and $\boldsymbol{A}_{i,(j)}$ are of size $(p_1+1)\times(p_1+1)$; the matrix $\mathbf{B}_{(j)}$ is of size $(p_1+1)\times(q_1+q_2)$; $\boldsymbol{\beta}_{\boldsymbol{r}_1,(j)}^0$ is of size $q_1\times 1$; $\boldsymbol{\beta}_{\boldsymbol{x}_i}^0$ is of size $p_1\times 1$, possibly equal to zero if the SE has less lags of $\boldsymbol{x}_t$ that the RF (i.e. if $\tilde{q}_1 < \tilde{p}$, $\boldsymbol{\beta}_{\boldsymbol{x}_i}^0 = 0$ for $i > \tilde{q}_1$); $\boldsymbol{\delta}_{y_i,(j)}^0$ is of size $p_1\times 1$; $\boldsymbol{\Delta}_{\boldsymbol{x}_i,(j)}^{0\prime}$ is of size $p_1\times p_1$; $\beta_{y_i}^0$ is a scalar parameter, possibly equal to zero if the number of lags is $\tilde{p}_1 < \tilde{p}$ (i.e. if $\tilde{p}_1 < \tilde{p}$ then $\beta_{y_i}^0 = 0$ for $i > \tilde{p}_1$), and $\mathbf{0}_{q_2}$ is a $q_2\times 1$ vector of zeros. Denote

$$\boldsymbol{C}_i(t,T) = \boldsymbol{A}_0^{-1}\boldsymbol{A}_i(t,T) = \boldsymbol{A}_0^{-1}\sum_{j=1}^{h+1}\boldsymbol{A}_{i,(j)}1_{t\in I_j^*}, \ \boldsymbol{J}(t,T) = \boldsymbol{A}_0^{-1}\mathbf{B}(t,T), \text{ and } \boldsymbol{e}_t = \boldsymbol{A}_0^{-1}\boldsymbol{\varepsilon}_t. \tag{64}$$

Then, (62) can be expressed as in (23) (see Section 3).

By Assumption 5, $\boldsymbol{\Delta}_{(j)}^0 = \boldsymbol{\Delta}^0 + O(T^{-\rho})$, hence $\boldsymbol{A}_{i,(j)} = \boldsymbol{A}_i + O(T^{-\rho})$, $\mathbf{B}_{(j)} = \mathbf{B} + O(T^{-\rho})$, $\boldsymbol{C}_{i,(j)} = \boldsymbol{C}_i + O(T^{-\rho})$, $\boldsymbol{J}_{(j)} = \boldsymbol{J} + O(T^{-\rho})$, $j = 1, \ldots, h+1$, where $\boldsymbol{\Delta}^0$, $\boldsymbol{A}_i$, $\mathbf{B}$, $\boldsymbol{C}_i$ and $\boldsymbol{J}$ are taken to be the common limiting values of $\boldsymbol{\Delta}_{(j)}^0$, $\boldsymbol{A}_{i,(j)}$, $\mathbf{B}_{(j)}$, $\boldsymbol{C}_{i,(j)}$ and $\boldsymbol{J}_{(j)}$ respectively, for all segments I_j^*.

Case (II). When there is a break in SE ($m = 1$) and when there are breaks in RF ($h > 0$), assume $\pi_{k-1}^0 < \lambda_1^0 \leq \pi_k^0$, $k = 1, \ldots, h+1$. Define $\ell = h$ if $\lambda_1^0 = \pi_k^0$; $\ell = h+1$ if $\lambda_1^0 < \pi_k^0$. Then,

$$\boldsymbol{A}_0(t,T)\tilde{\boldsymbol{y}}_t = \sum_{i=1}^{\tilde{p}} \boldsymbol{A}_i(t,T)\tilde{\boldsymbol{y}}_{t-i} + \mathbf{B}(t,T)\boldsymbol{r}_t + \boldsymbol{\varepsilon}_t, \tag{65}$$

with $\boldsymbol{A}_0(t,T) = \sum_{j=1}^{2}\boldsymbol{A}_{0,(j)}1_{t\in I_j^0}$, $\boldsymbol{A}_i(t,T) = \sum_{j=1}^{\ell+1}\boldsymbol{A}_{i,(j)}1_{t\in I_j^{*0}}$ and $\mathbf{B}(t,T) = \sum_{j=1}^{\ell+1}\mathbf{B}_{(j)}1_{t\in I_j^{*0}}$, where $I_j^0 = (T_{j-1}^0 + 1, \ldots, T_j^0)$, and if $\ell = h$ (the break in SE coincides with one of the breaks in RF), then $I_j^{*0} = I_j^* = (T_{j-1}^* + 1, \ldots, T_j^*)$, for $j = 1, 2, \ldots, \ell+1$. If $\ell = h+1$ (the break in SE does not coincide with any of the breaks in RF), then $I_j^{*0} = I_j^* = (T_{j-1}^* + 1, \ldots, T_j^*)$ if $j \leq k-1$, $I_k^{*0} = (T_{k-1}^*, \ldots, T_1^0) = ([T\pi_{k-1}^0], \ldots, [T\lambda_1^0])$, $I_{k+1}^{*0} = (T_1^0 + 1, \ldots, T_k^*) = ([T\lambda_1^0] + 1, \ldots, [T\pi_k^0])$, $I_j^{*0} = (T_{j-2}^* + 1, \ldots, T_{j-1}^*)$, $j = k+2, \ldots, \ell+1$, so that

$$\boldsymbol{A}_{0,(j)} = \begin{pmatrix} 1 & -\boldsymbol{\beta}_{\boldsymbol{x},(j)}^{0\prime} \\ \mathbf{0}_{p_1} & \boldsymbol{I}_{p_1} \end{pmatrix}. \tag{66}$$

Then if $\ell = h$:

$$\boldsymbol{A}_{i,(j)} = \begin{cases} \begin{pmatrix} \beta_{y_i,(1)}^0 & \boldsymbol{\beta}_{\boldsymbol{x}_i,(1)}^0 \\ \boldsymbol{\delta}_{y_i,(j)}^0 & \boldsymbol{\Delta}_{\boldsymbol{x}_i,(j)}^{0\prime} \end{pmatrix}, & \text{for } j < k, \\ \\ \begin{pmatrix} \beta_{y_i,(2)}^0 & \boldsymbol{\beta}_{\boldsymbol{x}_i,(2)}^0 \\ \boldsymbol{\delta}_{y_i,(j)}^0 & \boldsymbol{\Delta}_{\boldsymbol{x}_i,(j)}^{0\prime} \end{pmatrix}, & \text{for } j \geq k, \end{cases} \text{ and } \mathbf{B}_{(j)} = \begin{cases} \begin{pmatrix} \boldsymbol{\beta}_{\boldsymbol{r}_1,(1)}^{0\prime} & \mathbf{0}_{q_2}' \\ \boldsymbol{\Delta}_{\boldsymbol{r}_1,(j)}^{0\prime} & \boldsymbol{\Delta}_{\boldsymbol{r}_2,(j)}^{0\prime} \end{pmatrix}, & \text{for } j < k, \\ \\ \begin{pmatrix} \boldsymbol{\beta}_{\boldsymbol{r}_1,(2)}^{0\prime} & \mathbf{0}_{q_2}' \\ \boldsymbol{\Delta}_{\boldsymbol{r}_1,(j)}^{0\prime} & \boldsymbol{\Delta}_{\boldsymbol{r}_2,(j)}^{0\prime} \end{pmatrix}, & \text{for } j \geq k, \end{cases} \tag{67}$$

[27] We set the number of lags for $\boldsymbol{r}_t$ to zero, but $\boldsymbol{r}_t$ could involve lags at the expense of additional notation.

but if $\ell = h+1$:

$$A_{i,(j)} = \begin{cases} \begin{pmatrix} \beta^0_{y_i,(1)} & \boldsymbol{\beta}^0_{\boldsymbol{x}_i,(1)} \\ \boldsymbol{\delta}^0_{y_i,(j)} & \boldsymbol{\Delta}^{0'}_{\boldsymbol{x}_i,(j)} \end{pmatrix}, & \text{for } j \leq k, \\ \begin{pmatrix} \beta^0_{y_i,(2)} & \boldsymbol{\beta}^0_{\boldsymbol{x}_i,(2)} \\ \boldsymbol{\delta}^0_{y_i,(k)} & \boldsymbol{\Delta}^{0'}_{\boldsymbol{x}_i,(k)} \end{pmatrix}, & \text{for } j = k+1, \\ \begin{pmatrix} \beta^0_{y_i,(2)} & \boldsymbol{\beta}^0_{\boldsymbol{x}_i,(2)} \\ \boldsymbol{\delta}^0_{y_i,(j-1)} & \boldsymbol{\Delta}^{0'}_{\boldsymbol{x}_i,(j-1)} \end{pmatrix}, & \text{for } j > k+1, \end{cases} \text{ and } \mathbf{B}_{(j)} = \begin{cases} \begin{pmatrix} \boldsymbol{\beta}^{0'}_{\boldsymbol{r}_1,(1)} & \mathbf{0}'_{q_2} \\ \boldsymbol{\Delta}^{0'}_{\boldsymbol{r}_1,(j)} & \boldsymbol{\Delta}^{0'}_{\boldsymbol{r}_2,(j)} \end{pmatrix}, & \text{for } j \leq k, \\ \begin{pmatrix} \boldsymbol{\beta}^{0'}_{\boldsymbol{r}_1,(2)} & \mathbf{0}'_{q_2} \\ \boldsymbol{\Delta}^{0'}_{\boldsymbol{r}_1,(k)} & \boldsymbol{\Delta}^{0'}_{\boldsymbol{r}_2,(k)} \end{pmatrix}, & \text{for } j = k+1, \\ \begin{pmatrix} \boldsymbol{\beta}^{0'}_{\boldsymbol{r}_1,(2)} & \mathbf{0}'_{q_2} \\ \boldsymbol{\Delta}^{0'}_{\boldsymbol{r}_1,(j-1)} & \boldsymbol{\Delta}^{0'}_{\boldsymbol{r}_2,(j-1)} \end{pmatrix}, & \text{for } j > k+1. \end{cases}$$

Then,

$$\boldsymbol{C}_i(t,T) = \left\{\sum_{j=1}^{2} \boldsymbol{A}_{0,(j)} 1_{t\in I^0_j}\right\}^{-1} \sum_{j=1}^{\ell+1} \boldsymbol{A}_{i,(j)} 1_{t\in I^{*0}_j}, \ \boldsymbol{J}(t,T) = \left\{\sum_{j=1}^{2} \boldsymbol{A}_{0,(j)} 1_{t\in I^0_j}\right\}^{-1} \sum_{j=1}^{\ell+1} \mathbf{B}_{(j)} 1_{t\in I^{*0}_j}, \tag{68}$$

$$\boldsymbol{e}_t = \sum_{j=1}^{2} \boldsymbol{A}^{-1}_{0,(j)} 1_{t\in I^0_j} \boldsymbol{\varepsilon}_t, \tag{69}$$

where we use the same notation $\boldsymbol{C}_i(t,T)$, $\boldsymbol{J}(t,T)$, $\boldsymbol{e}_t$ as in **Case (I)** for the simplicity of notation, but these matrices and vector are different (compare (64) to (68) and (69)). Then (65) can be expressed as in (23) (see Section 3).

By Assumptions 3 and 5, $\boldsymbol{\beta}^0_{(i)} = \boldsymbol{\beta}^0 + O(T^{-\alpha})$, $\boldsymbol{\Delta}^0_{(j)} = \boldsymbol{\Delta}^0 + O(T^{-\rho})$, for $j = 1,\ldots,h+1$, and $i = 1,\ldots,m+1$, as these assumptions imply that the limiting coefficients will be the same regardless of breaks. Hence, if $\ell = h$ or $\ell = h+1$, $\boldsymbol{A}_{0,(j)} = \boldsymbol{A}_0 + O_p(T^{-\alpha})$, $\boldsymbol{A}_{i,(j)} = \boldsymbol{A}_i + O(T^{-\gamma})$, $\mathbf{B}_{(j)} = \mathbf{B} + O(T^{-\gamma})$, $\boldsymbol{C}_{i,(j)} = \boldsymbol{C}_i + O(T^{-\gamma})$, $\boldsymbol{J}_{(j)} = \boldsymbol{J} + O(T^{-\gamma})$, $j = 1,\ldots,\ell+1$, where $\gamma = \rho$ if I^{*0}_j does not contain the break in SE, and $\gamma = \min(\alpha,\rho)$ if I^{*0}_j contains the break in SE; see Assumptions 3 and 5.

Under Assumption 7 we have the vector VMAX representation of $\tilde{\boldsymbol{y}}_t$:

$$\tilde{\boldsymbol{y}}_t = \sum_{l=0}^{\infty} \boldsymbol{H}_l(t,T)\boldsymbol{J}(t,T)\boldsymbol{r}_{t-l} + \sum_{l=0}^{\infty} \boldsymbol{H}_l(t,T)\boldsymbol{e}_{t-l}, \text{ where } \sum_{l=0}^{\infty} \|\boldsymbol{H}_l(t,T)\| < \infty, \tag{70}$$

where $\boldsymbol{e}_t = \boldsymbol{A}_0^{-1}\boldsymbol{\varepsilon}_t$ and $\boldsymbol{H}_l(t,T) = \sum_{i=1}^{h+1} 1_{t\in I^*_i} \boldsymbol{H}_{l,(i)}$ for **Case (I)**; $\boldsymbol{e}_t = \sum_{j=1}^{2} \boldsymbol{A}^{-1}_{0,(j)}\boldsymbol{\varepsilon}_t 1_{t\in I^0_j}$ and $\boldsymbol{H}_l(t,T) = \sum_{i=1}^{\ell+1} 1_{t\in I^{*0}_i} \boldsymbol{H}_{l,(i)}$ for **Case (II)**.

Denote $\boldsymbol{H}(L,t,T) = \sum_{l=0}^{\infty} \boldsymbol{H}_l(t,T)L^l$ and let $\boldsymbol{C}(L,t,T) = \boldsymbol{I}_{p_1+1} - \boldsymbol{C}_1(t,T)L - \ldots - \boldsymbol{C}_{\tilde{p}}(t,T)L^{\tilde{p}}$, where L is the backshift operator, $L\tilde{\boldsymbol{y}}_t = \tilde{\boldsymbol{y}}_{t-1}$. The explicit form of the $\boldsymbol{H}_l(t,T)$'s is obtained using the fact that $\boldsymbol{H}(L,t,T)\boldsymbol{C}(L,t,T) = \boldsymbol{I}_{p_1+1}$. From Lutkepohl (2007) p.22. it follows that the $\boldsymbol{H}_l(t,T)$'s can be computed recursively as $\boldsymbol{H}_l(t,T) = \sum_{i=1}^{\min(l,\tilde{p})} \boldsymbol{H}_{l-i}(t,T)\boldsymbol{C}_i(t,T)$, with $\boldsymbol{H}_0(t,T) = \boldsymbol{I}_{p_1+1}$ and $\boldsymbol{H}_l(t,T) = 0$ for $l < 0$, $t = 1,\ldots,T$. For example, suppose $\tilde{p} = 1$, then we have the VARX(1,0) model: $\tilde{\boldsymbol{y}}_t = \boldsymbol{C}_1(t,T)\tilde{\boldsymbol{y}}_{t-1} + \boldsymbol{J}(t,T)\boldsymbol{r}_t + \boldsymbol{e}_t$ with matrices of coefficients in the VMAX representation given by: $\boldsymbol{H}_0(t,T) = \boldsymbol{I}_{p_1+1}$, $\boldsymbol{H}_1(t,T) = \boldsymbol{H}_0(t,T)\boldsymbol{C}_1(t,T) = \boldsymbol{C}_1(t,T)$, $\boldsymbol{H}_2(t,T) = \boldsymbol{H}_1(t,T)\boldsymbol{C}_1(t,T) = \boldsymbol{C}_1(t,T)^2, \ldots$. If $\tilde{p} = 2$, then we have the VARX(2,0) model: $\tilde{\boldsymbol{y}}_t = \boldsymbol{C}_1(t,T)\tilde{\boldsymbol{y}}_{t-1} + \boldsymbol{C}_2(t,T)\tilde{\boldsymbol{y}}_{t-2} + \boldsymbol{J}(t,T)\boldsymbol{r}_t + \boldsymbol{e}_t$ with matrices of coefficients in the VMAX representation given by: $\boldsymbol{H}_0(t,T) = \boldsymbol{I}_{p_1+1}$, $\boldsymbol{H}_1(t,T) = \boldsymbol{C}_1(t,T)$, $\boldsymbol{H}_2(t,T) = \boldsymbol{H}_1(t,T)\boldsymbol{C}_1(t,T) + \boldsymbol{H}_0(t,T)\boldsymbol{C}_2(t,T) = \boldsymbol{C}_1(t,T)^2 + \boldsymbol{C}_2(t,T)$, $\boldsymbol{H}_3(t,T) = \boldsymbol{H}_2(t,T)\boldsymbol{C}_1(t,T) + \boldsymbol{H}_1(t,T)\boldsymbol{C}_2(t,T) = \boldsymbol{C}_1(t,T)^3 + \boldsymbol{C}_1(t,T)\boldsymbol{C}_2(t,T), \ldots$. We emphasize again that $\boldsymbol{H}_l(t,T)$ of size $(p_1+1)\times(p_1+1)$ is different for **Cases (I)** and **(II)**, and as seen above it is a nonlinear function of the parameters $\boldsymbol{C}_i(t,T)$, $i = 1,\ldots,\tilde{p}$. For **Case (I)**, the $\boldsymbol{C}_i(t,T)$'s are given in (64), and for **Case (II)** they

are given in (68). Denote $\tilde{\tilde{\boldsymbol{y}}}_t = (\tilde{\boldsymbol{y}}'_{t-1}, \ldots, \tilde{\boldsymbol{y}}'_{t-\tilde{p}})'$ (as defined in (5)) and let

$$\tilde{\boldsymbol{H}}_l(t,T) = \begin{pmatrix} \boldsymbol{H}_{l-1}(t-1,T) \\ \vdots \\ \boldsymbol{H}_{l-\tilde{p}}(t-\tilde{p},T) \end{pmatrix}, \qquad \tilde{\boldsymbol{J}}_l(t,T) = \begin{pmatrix} \boldsymbol{H}_{l-1}(t-1,T)\boldsymbol{J}(t-1,T) \\ \vdots \\ \boldsymbol{H}_{l-\tilde{p}}(t-\tilde{p},T)\boldsymbol{J}(t-\tilde{p},T) \end{pmatrix} \tag{71}$$

the matrices of coefficients of size $\tilde{p}(p_1+1)\times(p_1+1)$, for $l = 1, 2, \ldots$, where recall that $\boldsymbol{H}_l(t,T) = 0$ for $l < 0$, and $\tilde{\boldsymbol{J}}_l(t,T) = \tilde{\boldsymbol{H}}_l(t,T)\boldsymbol{J}(t,T)$. By Assumption 7, $\sum_{l=0}^{\infty} \|\boldsymbol{H}_l(t,T)\| < \infty$ which implies that $\sum_{l=0}^{\infty} \left\|\tilde{\boldsymbol{H}}_l(t,T)\right\| < \infty$, hence the VMAX representation of $\tilde{\tilde{\boldsymbol{y}}}_t$ is:

$$\tilde{\tilde{\boldsymbol{y}}}_t = \sum_{l=0}^{\infty} \tilde{\boldsymbol{J}}_l(t,T)\boldsymbol{r}_{t-l} + \sum_{l=0}^{\infty} \tilde{\boldsymbol{H}}_l(t,T)\boldsymbol{e}_{t-l}. \tag{72}$$

Assuming $\tilde{\boldsymbol{y}}_0 = \boldsymbol{0}_{p_1+1}$, we can express (23) for $t \geq 2$, for both **Cases (I)** and **(II)** as:

$$\tilde{\boldsymbol{y}}_t = \sum_{l=0}^{t-1} \boldsymbol{H}_l(t,T)\boldsymbol{J}(t,T)\boldsymbol{r}_{t-l} + \sum_{l=0}^{t-1} \boldsymbol{H}_l(t,T)\boldsymbol{e}_{t-l}. \tag{73}$$

Assuming $\tilde{\tilde{\boldsymbol{y}}}_1 = (\tilde{\boldsymbol{y}}'_0, \tilde{\boldsymbol{y}}'_{-1}, \ldots, \tilde{\boldsymbol{y}}'_{-\tilde{p}+1})' = \boldsymbol{0}_{\tilde{p}(p_1+1)}$, for $t \geq 2$,

$$\tilde{\tilde{\boldsymbol{y}}}_t = \sum_{j=1}^{t-1} \tilde{\boldsymbol{J}}_l(t,T)\boldsymbol{r}_{t-l} + \sum_{j=1}^{t-1} \tilde{\boldsymbol{H}}_l(t,T)\boldsymbol{e}_{t-l}. \tag{74}$$

For **Case (I)** denote by $\boldsymbol{H}_l$ the common limiting value of $\boldsymbol{H}_{l,(i)}$ for any segment I_i^*, $i = 1, \ldots, h$. For convenience of notation denote also by $\boldsymbol{H}_l$ in **Case (II)** the common limiting value of $\boldsymbol{H}_{l,(i)}$ for any segment I_i^{*0}, $i = 1, \ldots, \ell$. By considering first **Case (I)**, notice that for the VARX$(\tilde{p}, 0)$, for any integer $\tilde{p} \geq 1$, we have $\boldsymbol{H}_1(t,T) = \boldsymbol{C}_1(t,T) = \boldsymbol{C}_1 + O(T^{-\rho}) = \boldsymbol{H}_1 + O(T^{-\rho})$, where $\boldsymbol{H}_1$ is the common limiting value of $\boldsymbol{H}_{1,(i)}$, $i = 1, \ldots, h$. Moreover $\boldsymbol{H}_2(t,T) = \boldsymbol{H}_1(t,T)\boldsymbol{C}_1(t,T) = (\boldsymbol{H}_1 + O(T^{-\rho}))(\boldsymbol{C}_1 + O(T^{-\rho})) = \boldsymbol{H}_1\boldsymbol{C}_1 + O(T^{-\rho}) = \boldsymbol{H}_2 + O(T^{-\rho})$ for $\tilde{p} = 1$, while for $\tilde{p} > 2$, $\boldsymbol{H}_2(t,T) = \boldsymbol{H}_1(t,T)\boldsymbol{C}_1(t,T) + \boldsymbol{H}_0(t,T)\boldsymbol{C}_2(t,T) = \boldsymbol{H}_1\boldsymbol{C}_1 + \boldsymbol{C}_2 + O(T^{-\rho}) = \boldsymbol{H}_2 + O(T^{-\rho})$, where $\boldsymbol{H}_2$ is the common limiting value of $\boldsymbol{H}_{2,(i)}$, $i = 1, \ldots, h$. For $\tilde{p} = 1$, $\boldsymbol{H}_3(t,T) = \boldsymbol{H}_2(t,T)\boldsymbol{C}_1(t,T) = \boldsymbol{H}_2\boldsymbol{C}_1 + O(T^{-\rho}) = \boldsymbol{H}_3 + O(T^{-\rho})$; for $\tilde{p} = 2$, $\boldsymbol{H}_3(t,T) = \boldsymbol{H}_2(t,T)\boldsymbol{C}_1(t,T) + \boldsymbol{H}_1(t,T)\boldsymbol{C}_2(t,T) = \boldsymbol{H}_2\boldsymbol{C}_1 + \boldsymbol{H}_1\boldsymbol{C}_2 + O(T^{-\rho})$; for $\tilde{p} > 2$, $\boldsymbol{H}_3(t,T) = \boldsymbol{H}_2(t,T)\boldsymbol{C}_1(t,T) + \boldsymbol{H}_1(t,T)\boldsymbol{C}_2(t,T) + \boldsymbol{H}_0(t,T)\boldsymbol{C}_1(t,T) = \boldsymbol{H}_2\boldsymbol{C}_1 + \boldsymbol{H}_1\boldsymbol{C}_2 + \boldsymbol{C}_1 + O(T^{-\rho}) = \boldsymbol{H}_3 + O(T^{-\rho})$, where $\boldsymbol{H}_3$ is the common limiting value of $\boldsymbol{H}_{3,(i)}$, $i = 1, \ldots, h$. It can be shown using mathematical induction that $\boldsymbol{H}_l(t,T) = \boldsymbol{H}_l + O(T^{-\rho})$ for any integer $\tilde{p} > 0$, where $\boldsymbol{H}_l$ is the common limiting value of $\boldsymbol{H}_{l,(i)}$, $i = 1, \ldots, h+1$. Similarly for **Case (II)** we can show that $\boldsymbol{H}_l(t,T) = \boldsymbol{H}_l + O(T^{-\gamma})$, where $\boldsymbol{H}_l$ is the common limiting value of $\boldsymbol{H}_{l,(i)}$, $i = 1, \ldots, \ell+1$. Thus we conclude that for **Case (I)**, $\tilde{\boldsymbol{H}}_l(t,T) = \tilde{\boldsymbol{H}}_l + O(T^{-\rho})$, where $\tilde{\boldsymbol{H}}_l$ is the common limiting value of $\tilde{\boldsymbol{H}}_{l,(i)}$, $i = 1, \ldots, h+1$, while for **Case (II)**, $\tilde{\boldsymbol{H}}_l(t,T) = \tilde{\boldsymbol{H}}_l + O(T^{-\gamma})$, where $\tilde{\boldsymbol{H}}_l$ is the common limiting value of $\tilde{\boldsymbol{H}}_{l,(i)}$, $i = 1, \ldots, \ell+1$. Similarly, we conclude that for **Case (I)**, $\tilde{\boldsymbol{J}}_l(t,T) = \tilde{\boldsymbol{J}}_l + O(T^{-\rho})$, where $\tilde{\boldsymbol{J}}_l$ is the common limiting value of $\tilde{\boldsymbol{J}}_{l,(i)}$, $i = 1, \ldots, h+1$, while for **Case (II)**, $\tilde{\boldsymbol{J}}_l(t,T) = \tilde{\boldsymbol{J}}_l + O(T^{-\gamma})$, where $\tilde{\boldsymbol{J}}_l$ is the common limiting value of $\tilde{\boldsymbol{J}}_{l,(i)}$, $i = 1, \ldots, \ell+1$, and:

$$\tilde{\boldsymbol{H}}_{l,(i)} = \begin{pmatrix} \boldsymbol{H}_{l-1,(i)} \\ \vdots \\ \boldsymbol{H}_{l-\tilde{p},(i)} \end{pmatrix}, \qquad \tilde{\boldsymbol{J}}_{l,(i)} = \begin{pmatrix} \boldsymbol{H}_{l-1,(i)}\boldsymbol{J}_{(i)} \\ \vdots \\ \boldsymbol{H}_{l-\tilde{p},(i)}\boldsymbol{J}_{(i)} \end{pmatrix}, \tag{75}$$

with $i = 1, \ldots, h+1$ for **Case (I)** and $i = 1, \ldots, \ell+1$ for **Case (II)**.

B Appendix: Definitions for Section 4

Definition B1. *$\boldsymbol{P}(\boldsymbol{\lambda}, \boldsymbol{\pi}^0)$ is a $(k+1)\times(h+1)$ matrix, with $k=1,2$, which is equal to $(\lambda_1,\ldots,\lambda_{k+1})'$ if $h=0$, and otherwise it is equal to:*

$$\begin{pmatrix} \frac{\pi_1^0\wedge\lambda_1}{\Delta\pi_1^0} & \left[\frac{\pi_2^0\wedge\lambda_1-\pi_1^0}{\Delta\pi_2^0}\right]_+ & \left[\frac{\pi_3^0\wedge\lambda_1-\pi_2^0}{\Delta\pi_3^0}\right]_+ & \cdots & \left[\frac{\pi_h^0\wedge\lambda_1-\pi_{h-1}^0}{\Delta\pi_h^0}\right]_+ & \frac{\lambda_1-\pi_h^0\wedge\lambda_1}{\Delta\pi_{h+1}^0} \\ \left[\frac{\pi_1^0\wedge\lambda_2-\lambda_1}{\Delta\pi_1^0}\right]_+ & \left[\frac{\pi_2^0\wedge\lambda_2-\pi_1^0\vee\lambda_1}{\Delta\pi_2^0}\right]_+ & \left[\frac{\pi_3^0\wedge\lambda_2-\pi_2^0\vee\lambda_1}{\Delta\pi_3^0}\right]_+ & \cdots & \left[\frac{\pi_h^0\wedge\lambda_2-\pi_{h-1}^0\vee\lambda_1}{\Delta\pi_h^0}\right]_+ & \left[\frac{\lambda_2-\lambda_1\vee\pi_h^0}{\Delta\pi_{h+1}^0}\right]_+ \\ \left[\frac{\pi_1^0-\lambda_2}{\Delta\pi_1^0}\right]_+ & \left[\frac{\pi_2^0-\pi_1^0\vee\lambda_2}{\Delta\pi_2^0}\right]_+ & \left[\frac{\pi_3^0-\pi_2^0\vee\lambda_2}{\Delta\pi_3^0}\right]_+ & \cdots & \left[\frac{\pi_h^0-\pi_{h-1}^0\vee\lambda_2}{\Delta\pi_h^0}\right]_+ & \frac{1-\pi_h^0\vee\lambda_2}{\Delta\pi_{h+1}^0} \end{pmatrix},$$

where $\vee$ stands for maximum, and $\wedge$ for minimum, $[a]_+ = a\vee 0$ stands for the positive values of a, and $\Delta\pi_i^0 = \pi_i^0 - \pi_{i-1}^0$, $i=1,\ldots,h+1$. For example, if $k=1$ and $h=1$, then $\lambda_2=\pi_2^0=1$; letting $1_{\{\cdot\}}$ be the indicator function, we have:

$$\boldsymbol{P}(\boldsymbol{\lambda},\boldsymbol{\pi}^0) = \begin{pmatrix} \frac{\pi_1^0\wedge\lambda_1}{\pi_1^0} & 1_{\pi_1^0\le\lambda_1}\frac{\lambda_1-\pi_1^0}{1-\pi_1^0} \\ 1_{\pi_1^0>\lambda_1}\frac{\pi_1^0-\lambda_1}{\pi_1^0} & \frac{\pi_2^0-\pi_1^0\vee\lambda_1}{1-\pi_1^0} \end{pmatrix}.$$

Definition B2. *Let $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_u + \oplus_{q\times q}\left(\boldsymbol{\Sigma}_{\boldsymbol{v}}\odot\left((\boldsymbol{\beta}_{\boldsymbol{x}}^0\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime})\otimes\boldsymbol{E}_q\right)\right) + 2\oplus_{q\times q}\left(\boldsymbol{\Sigma}_{u\boldsymbol{v}}\odot\left((\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime}\otimes\boldsymbol{E}_q\right)\right)$, a $q\times q$ matrix, where $\oplus_{q\times q}$ is the operator that adds all $q\times q$ blocks of a $p_1q\times p_1q$ matrix, $\odot$ is the operator that multiplies element-wise two matrices of the same size, $\boldsymbol{E}_q$ is the $q\times q$ matrix of ones, and $\boldsymbol{\Sigma}_u, \boldsymbol{\Sigma}_{\boldsymbol{v}}, \boldsymbol{\Sigma}_{u\boldsymbol{v}}$ are defined in Assumption 13(i). Let $\tilde{\boldsymbol{\Sigma}}_1 = \tilde{\boldsymbol{\Sigma}}_u + \oplus_{q\times q}\left(\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}}\odot\left((\boldsymbol{\beta}_{\boldsymbol{x}}^0\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime})\otimes\boldsymbol{E}_q\right)\right) + 2\oplus_{q\times q}\left(\tilde{\boldsymbol{\Sigma}}_{u\boldsymbol{v}}\odot\left((\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime}\otimes\boldsymbol{E}_q\right)\right)$, where $\tilde{\boldsymbol{\Sigma}}_u, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}}, \tilde{\boldsymbol{\Sigma}}_{u\boldsymbol{v}}$ are defined in Assumption 13(ii). Let $\tilde{\tilde{\boldsymbol{\Sigma}}}_1 = \tilde{\tilde{\boldsymbol{\Sigma}}}_u + \oplus_{q\times q}\left(\tilde{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{v}}\odot\left((\boldsymbol{\beta}_{\boldsymbol{x}}^0\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime})\otimes\boldsymbol{E}_q\right)\right) + 2\oplus_{q\times q}\left(\tilde{\tilde{\boldsymbol{\Sigma}}}_{u\boldsymbol{v}}\odot\left((\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime}\otimes\boldsymbol{E}_q\right)\right)$, where $\tilde{\tilde{\boldsymbol{\Sigma}}}_u$, $\tilde{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{v}}$ and $\tilde{\tilde{\boldsymbol{\Sigma}}}_{u\boldsymbol{v}}$ are the restricted versions of $\tilde{\boldsymbol{\Sigma}}_u$, $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}}$ and $\tilde{\boldsymbol{\Sigma}}_{u\boldsymbol{v}}$ when the errors u_t and $\boldsymbol{v}_t$, $t=1,\ldots,T$, are heteroskedastic but not autocorrelated (see Section 5.1.2).*

Definition B3. *Let $\eta_1^* = \eta_1\lambda_1^0$, and $\eta_2^* = \lambda_1^0 + \eta_2(1-\lambda_1^0)$, and, for any scalar a, let $[a]_+$ denote the positive part, i.e. $a = a\,1_{a>0}$. Then, for $h=0$, $\boldsymbol{P}(\eta_1;\boldsymbol{\pi}^0), \boldsymbol{P}(\eta_2;\boldsymbol{\pi}^0)$ are $2\times(h+1)$ matrices, equal to $(\eta_1^*, \lambda_1^0-\eta_1^*)'$ and $(\eta_2^*-\lambda_1^0, 1-\eta_2^*)'$. Otherwise, $\boldsymbol{P}(\eta_1;\boldsymbol{\pi}^0)$ is:*

$$\begin{pmatrix} \frac{\pi_1^0\wedge\eta_1^*}{\Delta\pi_1^0} & \left[\frac{\pi_2^0\wedge\eta_1^*-\pi_1^0}{\Delta\pi_2^0}\right]_+ & \left[\frac{\pi_3^0\wedge\eta_1^*-\pi_2^0}{\Delta\pi_3^0}\right]_+ & \cdots & \left[\frac{\pi_h^0\wedge\eta_1^*-\pi_{h-1}^0}{\Delta\pi_h^0}\right]_+ & \left[\frac{\eta_1^*-\pi_h^0}{\Delta\pi_{h+1}^0}\right]_+ \\ \left[\frac{\pi_1^0\wedge\lambda_1^0-\eta_1^*}{\Delta\pi_1^0}\right]_+ & \left[\frac{\pi_2^0\wedge\lambda_1^0-\pi_1^0\vee\eta_1^*}{\Delta\pi_2^0}\right]_+ & \left[\frac{\pi_3^0\wedge\lambda_1^0-\pi_2^0\vee\eta_1^*}{\Delta\pi_3^0}\right]_+ & \cdots & \left[\frac{\pi_h^0\wedge\lambda_1^0-\pi_{h-1}^0\vee\eta_1^*}{\Delta\pi_h^0}\right]_+ & \left[\frac{\lambda_1^0-\pi_h^0\vee\eta_1^*}{\Delta\pi_{h+1}^0}\right]_+ \end{pmatrix}$$

and $\boldsymbol{P}(\eta_2;\boldsymbol{\pi}^0)$ is:

$$\begin{pmatrix} \left[\frac{\pi_1^0\wedge\eta_2^*-\lambda_1^0}{\Delta\pi_1^0}\right]_+ & \left[\frac{\pi_2^0\wedge\eta_2^*-\pi_1^0\vee\lambda_1^0}{\Delta\pi_2^0}\right]_+ & \left[\frac{\pi_3^0\wedge\eta_2^*-\pi_2^0\vee\lambda_1^0}{\Delta\pi_3^0}\right]_+ & \cdots & \left[\frac{\pi_h^0\wedge\eta_2^*-\pi_{h-1}^0\vee\lambda_1^0}{\Delta\pi_h^0}\right]_+ & \left[\frac{\eta_2^*-\pi_h^0\vee\lambda_1^0}{\Delta\pi_{h+1}^0}\right]_+ \\ \left[\frac{\pi_1^0-\eta_2^*}{\Delta\pi_1^0}\right]_+ & \left[\frac{\pi_2^0-\pi_1^0\vee\eta_2^*}{\Delta\pi_2^0}\right]_+ & \left[\frac{\pi_3^0-\pi_2^0\vee\eta_2^*}{\Delta\pi_3^0}\right]_+ & \cdots & \left[\frac{\pi_h^0-\pi_{h-1}^0\vee\eta_2^*}{\Delta\pi_h^0}\right]_+ & \frac{1-\pi_h^0\vee\eta_2^*}{\Delta\pi_{h+1}^0} \end{pmatrix}.$$

For example, if $h=1$, then:

$$\boldsymbol{P}(\eta_1;\boldsymbol{\pi}^0) = \begin{pmatrix} \frac{\pi_1^0\wedge\eta_1^*}{\pi_1^0} & 1_{\pi_1^0\le\eta_1^*}\frac{\eta_1^*-\pi_1^0}{1-\pi_1^0} \\ 1_{\pi_1^0>\eta_1^*}\frac{\pi_1^0\wedge\lambda_1^0-\eta_1^*}{\pi_1^0} & 1_{\lambda_1^0>(\pi_1^0\vee\eta_1^*)}\frac{\lambda_1^0-\pi_1^0\vee\eta_1^*}{1-\pi_1^0} \end{pmatrix} \text{ and } \boldsymbol{P}(\eta_2;\boldsymbol{\pi}^0) = \begin{pmatrix} 1_{\lambda_1^0<(\pi_1^0\vee\eta_2^*)}\frac{\pi_1^0\wedge\eta_2^*-\lambda_1^0}{\pi_1^0} & 1_{\pi_1^0<\eta_2^*}\frac{\eta_2^*-\pi_1^0\vee\lambda_1^0}{1-\pi_1^0} \\ 1_{\pi_1^0>\eta_2^*}\frac{\pi_1^0-\eta_2^*}{\pi_1^0} & \frac{1-\pi_1^0\vee\eta_2^*}{1-\pi_1^0} \end{pmatrix}.$$

C Appendix: Tables for Section 6

Table 1: Rejection probabilities for testing H_0 : 0 breaks in SE, H_1 : 1 break in SE; sup-F test; DGP: 0 break in RF, 0 break in SE.

	BP asymptotic			IR bootstrap			IF bootstrap		
				Case (A)					
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.10	0.05	0.01	0.09	0.05	0.01	0.10	0.05	0.01
240	0.09	0.05	0.01	0.10	0.05	0.01	0.05	0.02	0.004
480	0.09	0.05	0.01	0.10	0.05	0.01	0.06	0.03	0.01
				Case (B)					
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.11	0.06	0.02	0.10	0.05	0.01	0.10	0.05	0.01
240	0.10	0.06	0.02	0.10	0.05	0.01	0.06	0.03	0.004
480	0.10	0.05	0.01	0.10	0.05	0.01	0.05	0.02	0.002

BP critical values: 3.57, 4.05, 5.06 for $0.10, 0.05, 0.01$ nominal levels

Table 2: Rejection probabilities for testing H_0 : 0 break in SE, H_1 : 1 break in SE; sup-F test; DGP: 1 break in RF, 0 breaks in SE.

	BP asymptotic			BCH asymptotic			IR bootstrap			IF bootstrap		
						Case (A)						
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.15	0.09	0.03	0.10	0.05	0.01	0.11	0.06	0.01	0.10	0.05	0.01
240	0.15	0.09	0.02	0.09	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
480	0.16	0.09	0.03	0.09	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
						Case (B)						
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.17	0.11	0.04	0.11	0.07	0.02	0.11	0.06	0.01	0.12	0.06	0.01
240	0.16	0.09	0.03	0.10	0.05	0.01	0.11	0.06	0.01	0.10	0.05	0.01
480	0.16	0.09	0.03	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01

BP critical values: 3.57, 4.05, 5.06, BCH critical values: 4.00, 4.53, 5.85.

Table 3: Rejection probabilities for testing H_0 : 1 break in SE, H_1 : 2 breaks in SE; sup-F test; DGP: 0 break in RF, 1 break in SE.

	BP asymptotic			IR bootstrap			IF bootstrap		
				Case (A)					
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.03	0.01	0.002	0.10	0.05	0.01	0.11	0.06	0.01
240	0.06	0.02	0.004	0.10	0.05	0.01	0.12	0.06	0.01
480	0.08	0.03	0.012	0.10	0.05	0.01	0.11	0.06	0.01
				Case (B)					
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.07	0.04	0.01	0.12	0.06	0.01	0.14	0.07	0.02
240	0.09	0.04	0.01	0.11	0.06	0.01	0.12	0.06	0.01
480	0.09	0.04	0.01	0.10	0.05	0.01	0.11	0.05	0.01

BP critical values: 16.11, 18.11, 21.97 for $0.10, 0.05, 0.01$ nominal levels

Table 4: Rejection probabilities for testing H_0 : 1 break in SE, H_1 : 2 breaks in SE; sup-F test; DGP: 1 break in RF, 1 break in SE.

	BP asymptotic			BCH asymptotic			IR bootstrap			IF bootstrap		
						Case (A)						
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.02	0.01	0.001	0.04	0.01	0.002	0.10	0.05	0.01	0.11	0.06	0.01
240	0.04	0.01	0.002	0.07	0.03	0.01	0.10	0.05	0.01	0.11	0.05	0.01
480	0.05	0.02	0.003	0.09	0.04	0.01	0.09	0.05	0.01	0.11	0.06	0.01
						Case (B)						
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.05	0.02	0.01	0.09	0.04	0.013	0.12	0.06	0.02	0.13	0.07	0.02
240	0.07	0.03	0.01	0.11	0.06	0.013	0.11	0.06	0.01	0.12	0.06	0.01
480	0.06	0.03	0.01	0.11	0.06	0.012	0.10	0.05	0.01	0.11	0.05	0.01

BP critical values: 16.11, 18.11, 21.97, BCH critical values: 14.76, 16.43, 20.24

Table 5: Rejection probabilities for testing H_0 : 0 break in SE, H_1 : 1 break in SE; sup-$Wald$ test; DGP: 0 break in RF, 0 break in SE.

	BP asymptotic			WR bootstrap			WF bootstrap		
				Case (A)					
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.43	0.34	0.19	0.10	0.05	0.01	0.10	0.05	0.01
240	0.26	0.18	0.07	0.10	0.05	0.01	0.10	0.05	0.01
480	0.18	0.10	0.03	0.10	0.05	0.01	0.10	0.05	0.01
				Case (C)					
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.43	0.33	0.19	0.10	0.05	0.01	0.10	0.05	0.01
240	0.25	0.17	0.07	0.10	0.05	0.01	0.10	0.05	0.01
480	0.17	0.10	0.03	0.11	0.05	0.01	0.10	0.05	0.01

BP critical values: 14.26, 16.19, 20.23

Table 6: Rejection probabilities for testing H_0 : 0 break in SE, H_1 : 1 break in SE; sup-$Wald$ test; DGP: 1 break in RF, 0 breaks in SE.

	BP asymptotic			BCH asymptotic			WR bootstrap			WF bootstrap		
						Case (A)						
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.52	0.41	0.25	0.36	0.27	0.13	0.12	0.06	0.01	0.11	0.06	0.01
240	0.35	0.25	0.12	0.20	0.13	0.04	0.11	0.06	0.01	0.11	0.06	0.01
480	0.25	0.16	0.06	0.12	0.07	0.01	0.11	0.06	0.01	0.11	0.06	0.01
						Case (C)						
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.52	0.42	0.25	0.37	0.28	0.12	0.12	0.06	0.01	0.11	0.05	0.01
240	0.34	0.24	0.12	0.20	0.12	0.04	0.11	0.06	0.01	0.11	0.05	0.01
480	0.25	0.16	0.06	0.13	0.07	0.01	0.10	0.06	0.01	0.10	0.06	0.01

BP critical values: 14.26, 16.19, 20.23, BCH critical values: 17.28, 19.71, 26.12

Table 7: Rejection probabilities for testing H_0 : 1 break in SE, H_1 : 2 breaks in SE; sup-$Wald$ test; DGP: 0 break in RF, 1 break in SE.

	BP asymptotic			WR bootstrap			WF bootstrap		
				Case (A)					
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.98	0.97	0.94	0.09	0.04	0.01	0.09	0.05	0.01
240	0.74	0.65	0.50	0.10	0.05	0.01	0.10	0.05	0.01
480	0.42	0.31	0.17	0.10	0.05	0.01	0.11	0.05	0.01
				Case (C)					
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.98	0.97	0.94	0.08	0.04	0.01	0.10	0.05	0.01
240	0.74	0.65	0.47	0.10	0.05	0.01	0.10	0.05	0.01
480	0.42	0.31	0.17	0.11	0.05	0.01	0.10	0.05	0.01

BP critical values: 16.11, 18.11, 21.97

Table 8: Rejection probabilities for testing H_0 : 1 break in SE, H_1 : 2 breaks in SE; sup-$Wald$ test; DGP: 1 break in RF, 1 break in SE.

	BP asymptotic			BCH asymptotic			WR bootstrap			WF bootstrap		
						Case (A)						
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.99	0.98	0.96	0.98	0.96	0.92	0.08	0.04	0.01	0.06	0.03	0.004
240	0.75	0.67	0.53	0.62	0.51	0.32	0.11	0.05	0.01	0.11	0.06	0.01
480	0.41	0.31	0.18	0.26	0.17	0.07	0.11	0.06	0.01	0.11	0.06	0.02
						Case (C)						
T	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
120	0.99	0.98	0.96	0.98	0.96	0.91	0.09	0.04	0.01	0.05	0.02	0.002
240	0.75	0.67	0.52	0.62	0.52	0.33	0.11	0.05	0.01	0.12	0.06	0.01
480	0.41	0.31	0.18	0.27	0.18	0.07	0.12	0.06	0.02	0.12	0.06	0.02

BP critical values: 16.11, 18.11, 21.97, BCH critical values: 19.21, 22.10, 29.23

References

Andrews, D. W. K. (1993). 'Tests for parameter instability and structural change with unknown change point', *Econometrica*, 61: 821–856.

Andrews, D. W. K., and Ploberger, W. (1994). 'Optimal tests when a nuisance parameter is present only under the alternative', *Econometrica*, 62: 1383–1414.

Antoshin, S., Berg, A., and Souto, M. (2008). 'Testing for Structural Breaks in Small Samples', *IMF Working Paper*.

Bai, J. (1997). 'Estimation of a change point in multiple regression models', *Review of Economics and Statistics*, 79: 551–563.

Bai, J., and Perron, P. (1998). 'Estimating and testing linear models with multiple structural changes', *Econometrica*, 66: 47–78.

——— (2003). 'Computation and Analysis of Multiple Structural Change Models', *Journal of Applied Econometrics*, 18: 1–22.

Banerjee, A., Lazarova, S., and Urga, G. (2002). 'Bootstrapping Structural Tests for Multiple Structural Breaks', *London Business School Discussion Papers*, 17.

Bekaert, G., and Hodrick, R. J. (2001). 'Expectations hypothesis tests', *Journal of Finance*, 56: 1357–1394.

Beran, R. (1986). 'Discussion of "Jackknife bootstrap and other resampling methods in regression analysis" by C. F. J. Wu', *Annals of Statistics*, 14: 1295–1298.

Berg, A., Ostry, J. D., and Zettelmeyer, J. (2012). 'What Makes Growth Sustained?', *Journal of Development Economics*, 98: 149–166.

Bergamelli, M., and Urga, G. (2013). 'Detecting Multiple Structural Breaks: Dummy Saturation vs Sequential Bootstrapping', *Discussion Paper, Cass Business School, London*.

Berkowitz, J., and Kilian, L. (2000). 'Recent developments in bootstrapping time series', *Econometric Reviews*, 19: 1–48.

Bollerslev, T. (1986). 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics*, 37: 307–327.

——— (1990). 'Modelling the coherence on hosrt-run nominal exchange rates: a multivariate GARCH model', *Review of Economics and Statistics*, 72: 498–505.

Cavaliere, G., Rahbek, A., and Taylor, A. R. (2010). 'Cointegrating rank testing under conditional heteroskedasticity', *Econometric Theory*, 26: 1719–1760.

Christiano, L. J. (1992). 'Searching for a Break in GNP', *Journal of Business and Economic Statistics*, 10(3): 237–250.

Clark, T. E. (2006). 'Disaggregate Evidence on the Persistence of Consumer Price Inflation', *Journal of Applied Econometrics*, 21: 563–587.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press, Oxford, UK.

Davidson, R. (2007). 'Bootstrapping Econometric Models', *Quantile*, 3: 13–36, http://russell-davidson.arts.mcgill.ca/articles/quantile.pdf.

Davidson, R., and Flachaire, E. (2008). 'The Wild Bootstrap, Tamed at Last', *Journal of Econometrics*, 146: 162–169.

Davidson, R., and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press.

——— (2010). 'Wild Bootstrap Tests for IV Regression', *Journal of Business and Economic Statistics*, 28: 128–144.

de Peretti, C., and Urga, G. (2004). 'Stopping tests in the sequential estimation for multiple structural breaks', *Working Paper Cass Business School*.

Diebold, F. X., and Chen, C. (1996). 'Testing structural stability with endogenous break-point: A size comparison of analytic and bootstrap procedures', *Journal of Econometrics*, 70: 221–241.

Efron, B. (1979). 'Bootstrap Methods: Another Look at the Jackknife', *Annals of Statistics*, 7(1): 1–26.

Feldstein, M., and Stock, J. H. (1994). 'The Use of Monetary Aggregate to Target Nominal GDP', in *Monetary Policy*. Chicago: University of Chicago Press.

Flachaire, E. (1999). 'A better way to bootstrap pairs', *Economics Letters*, 64: 257–262.

Gonçalves, S., and Kilian, L. (2004). 'Bootstrapping autoregressions with conditional heteroskedasticity of unknown form', *Journal of Econometrics*, 123: 89–120.

Hall, A. R., Han, S., and Boldea, O. (2012). 'Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors', *Journal of Econometrics*, 170(2): 281–302.

Hall, A. R., Osborn, D. R., and Sakkas, N. D. (2013). 'Inference about Structural Breaks using Information Criteria', *The Manchester School*, 81: 54–81.

Hansen, B. (2001). 'The New Econometrics of Structural Change: Dating Changes in U.S. Labor Productivity', *Journal of Economic Perspectives*, 15: 117–128.

Hansen, B. E. (2000). 'Testing for structural change in conditional models', *Journal of Econometrics*, 97: 93–115.

Hodrick, R. J. (1992). 'Dividend yields and expected stock returns: alternative procedures for inference and measurement', *Review of Financial Studies*, 5: 357–386.

Hsu, C.-C. (2005). 'Long memory or structural changes: An empirical examination on inflation rates', *Economics Letters*, 88(2): 289–294.

Kreiss, J. P. (1997). 'Asymptotic properties of the residual bootstrap for autoregressions', *Manuscript, Institute for Mathematical Stochastics, Technical University of Breunschweig, Germany*.

Ledoit, O., Santa-Clara, P., and Wolf, M. (2003). 'Flexible multivariate GARCH modelling with an application to international markets', *Review of Economics and Statistics*, 85: 735–747.

Levin, A. T., and Piger, J. M. (2006). 'Is Inflation Persistence Intrinsic in Industrial Economies', *Federal Reserve of St. Louis Working Paper*.

Liu, R. Y. (1988). 'Bootstrap procedure under some non-i.i.d. models', *Annals of Statistics*, 16: 1696–1708.

Lutkepohl, H. (2007). *New Introduction to Multiple Time Series AnalysisSpringer Verlag*.

Mammen, E. (1993). 'Bootstrap and Wild Bootstrap for High Dimensional Linear Models', *Annals of Statistics*, 21: 255–285.

Morana, C., and Beltratti, A. (2004). 'Structural change and long-range dependence in volatility of exchange rates: either, neither or both?', *Journal of Empirical Finance*, 11(5): 629–658.

O'Reilly, G., and Whelan, K. (2004). 'Has Euro-Area Inflation Persistence Changed Over Time?', *European Central Bank Working Paper*, (335).

Perron, P. (1989). 'The Great Crash, the Oil Price Shock and the Unit Root Hypothesis', *Econometrica*, 57: 1361–1401.

Perron, P., and Vogelsang, T. J. (1992). 'Nonstationarity and Level Shifts With an Application to Purchasing Power Parity', *Journal of Business and Economics Statistics*, 10(3): 301–320.

Qu, Z., and Perron, P. (2007). 'Estimating and testing structural changes in multivariate regressions', *Econometrica*, 75: 459–502.

Quintos, C., Fan, Z., and Phillips, P. C. B. (2001). 'Structural Change Tests in Tail Behaviour and the Asian Crisis', *Review of Economic Studies*, 68(633-663).

Stock, J., and Watson, M. (2005). 'Understanding Changes in International Business Cycle Dynamics', *Journal of the European Economic Association*, 3(5): 968–1006.

Stock, J. H. (1994). 'Unit Roots, Structural Breaks and Trends', in R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. IV. Elsevier.

Wooldridge, J., and White, H. (1988). 'Some invariance principles and central limit theorems for dependent heterogeneous processes', *Econometric Theory*, 4: 210–230.

Wu, C. F. J. (1986). 'Jackknife bootstrap and the other resampling methods in regression analysis', *Annals of Statistics*, 14: 1261–1295.