Quantitative Economics 11 (2020), 1019–1057 1759-7331/20201019

Testing jointly for structural changes in the error variance and coefficients of a linear regression model

Pierre Perron
Department of Economics, Boston University

Yohei Yamamoto
Department of Economics, Hitotsubashi University

Jing Zhou
Seeking Sense (Shanghai) Investment Management Co., Ltd.

We provide a comprehensive treatment for the problem of testing jointly for structural changes in both the regression coefficients and the variance of the errors in a single equation system involving stationary regressors. Our framework is quite general in that we allow for general mixing-type regressors and the assumptions on the errors are quite mild. Their distribution can be nonnormal and conditional heteroskedasticity is permitted. Extensions to the case with serially correlated errors are also treated. We provide the required tools to address the following testing problems, among others: (a) testing for given numbers of changes in regression coefficients and variance of the errors; (b) testing for some unknown number of changes within some prespecified maximum; (c) testing for changes in variance (regression coefficients) allowing for a given number of changes in the regression coefficients (variance); (d) a sequential procedure to estimate the number of changes present. These testing problems are important for practical applications as witnessed by interests in macroeconomics and finance where documenting structural changes in the variability of shocks to simple autoregressions or vector autoregressive models have been a concern.

Keywords. Change-point, variance shift, conditional heteroskedasticity, likelihood ratio tests.

JEL classification. C22.

1. Introduction

Both the statistics and econometrics literature contain a vast amount of work on issues related to structural changes with unknown break dates, most of it designed for a single change (for an extensive review, see Perron (2006) and Casini and Perron (2019b)). The

Pierre Perron: perron@bu.edu
Yohei Yamamoto: yohei.yamamoto@econ.hit-u.ac.jp
Jing Zhou: zhoujing112@foxmail.com
Perron acknowledges financial support from the National Science Foundation under Grant SES-0649350. We are grateful to Zhongjun Qu for comments and for pointing out an error in a previous draft.

 https://doi.org/10.3982/QE1332

problem of multiple structural changes has received attention mostly in the context of a single regression. Bai and Perron (1998, 2003a) provided a comprehensive treatment: consistency of estimates of the break dates, tests for structural changes, confidence intervals for the break dates, methods to select the number of breaks, and efficient algorithms to compute the estimates; see also Hawkins (1976). Perron and Qu (2006) extended this analysis to the case where arbitrary linear restrictions are imposed on the coefficients of the model. Also, Kurozumi and Tuvaandorj (2011) proposed an information criterion for the selection of the number of changes; see also Liu, Wu, and Zidek (1997). Bai, Lumsdaine, and Stock (1998) considered asymptotically valid inference for the estimate of a single break date in multivariate time series allowing stationary or integrated regressors as well as trends with estimation carried using a quasi maximum likelihood (QML) procedure. Also, Bai (2000) considered a segmented stationary VAR model estimated again by QML when the break can occur in the parameters of the conditional mean, the variance of the error term or both. Kejriwal and Perron (2008, 2010) dealt with multiple structural changes in a single equation cointegrated model. Perron and Yamamoto (2014) derived the limit distribution of the estimates of the break dates in models with endogenous regressors estimated via an instrumental variable method, while they argue in Perron and Yamamoto (2015) that using standard least-squares methods is preferable both for estimation and testing. Casini and Perron (2019a) provided a limit distribution of the least-squares estimate of the break date in a linear model based on a continuous-time asymptotic framework, which delivers substantial improvements with respect to inference using the concept of highest density regions.

With respect to testing for changes in the variance of the regression error, the results are quite sparse. Horváth (1993) considered a change in the mean and variance (occurring at the same time) of a sequence of i.i.d. random variables with moments corresponding to those of a normal distribution. Davis, Huang, and Yao (1995) extended the analysis to an autoregressive process under similar conditions. Aue, Hormann, Horváth, and Reimherr (2009) proposed nonparametric tests for changes in the variances or autocovariances of multivariate linear or nonlinear time series models. Deng and Perron (2008) extended the CUSUM of squares (or CUSQ) test of Brown, Durbin, and Evans (1975) allowing general conditions on the regressors and the errors (as suggested by Inclán and Tiao (1994) for normally distributed time series). Xu (2013) provided a further extension with a robust estimate of the long-run variance of the squared errors of closer relevance to our objectives. Andrews (1993) considered a one-time structural change under a Generalized Method of Moment (GMM) setting, thereby allowing for changes in both coefficients and variance though occurring at the same date; see McConnell and Pérez-Quirós (2000) for a related application. Qu and Perron (2007a) considered a multivariate system estimated by quasi maximum likelihood, which provides methods to estimate models with structural changes in both the regression coefficients and the covariance matrix of the errors. They provide a limit distribution theory for inference about the break dates and also consider testing for multiple structural changes, though restricted to normally distributed errors and breaks in coefficients and variance occurring at different dates.

We build on Qu and Perron (2007a) to provide a comprehensive treatment of testing jointly for structural changes in both the regression coefficients and the variance of the errors in a single equation involving stationary regressors, allowing the break dates to be different or overlap. Our framework is general and allows for general mixing-type regressors. The assumptions on the errors are mild; their distribution can be nonnormal and conditional heteroskedasticity is permitted. Extensions to the case with serially correlated errors are also treated. We provide the required tools to address the following testing problems, among others: (a) testing for given numbers of changes in regression coefficients and variance of the errors; (b) testing for some unknown number of changes within some prespecified maximum; (c) testing for changes in variance (regression coefficients) allowing for a given number of changes in the regression coefficients (variance); (d) sequential procedures to estimate the number of changes present. Note that we adopt a QML approach instead of one based on GMM. Either could be used in principle. The main advantage of using the QML approach based on normal errors is first that it allows a natural extension of Bai and Perron (1998) widely used in practice. Second, and more importantly perhaps, we can use the efficient algorithm developed in Qu and Perron (2007a). This is especially important in the current context since even only two breaks in coefficients and variance implies four possible break dates. Hence a computationally efficient method to estimate the break dates is needed.

These testing problems are important for practical applications; for example, documenting structural changes in the variability of shocks in autoregressive models; see Blanchard and Simon (2001), Herrera and Pesavento (2005), Kim and Nelson (1999), McConnell and Pérez-Quirós (2000), Sensier and van Dijk (2004), and Stock and Watson (2002). Given the lack of proper testing procedures, a common approach is to apply a sup-Wald type tests (e.g., Andrews (1993), Bai and Perron (1998)) for changes in the mean of the absolute value of the estimated residuals, a rather ad hoc procedure. To test for a change in variance only (imposing no change in the regression coefficients), one can apply a CUSUM of squares test to the estimated residuals, which is adequate only if no change in coefficient is present. Often, changes in both coefficients and variance occur at possibly different dates. A common method is to first test for changes in the regression coefficients and conditioning on the break dates found, then test for changes in variance. This is clearly inappropriate as in the first step the tests suffers for severe size distortions. Also, neglecting changes in regression coefficients when testing for changes in variance induces both size distortions and a loss of power; for example, Perron and Yamamoto (2019a) and Pitarakis (2004). Hence, what is needed is a joint approach. To do so, our testing procedures are based on quasi likelihood ratio tests using a likelihood function for identically and independently distributed normal errors. We then apply corrections to have limit distributions free of nuisance parameters with nonnormal distribution and conditional heteroskedasticity. We also consider extensions that allow for serial correlation.

The empirical usefulness of our proposed procedure is perhaps best explained via applications related to changes in the variance of many macroeconomic variables (i.e., the great moderation); see Gadea, Gómez-Loscos, and Pérez-Quirós (2018) and Perron and Yamamoto (2019b). The testing issues of interest are, among others: (a) testing for a change in variance in 1984 (the commonly accepted date for the start of the great moderation); (b) testing for an additional change in variance, say following the great recession of 2007; (c) estimating the total number of changes; (d) testing whether any changes are present; (e) performing all these tests allowing for changes in the parameters of a conditional regression model (e.g., a change in slope in 1973 for GDP as argued in Perron (1989)); (f) performing all the corresponding tests when testing for changes in the regression parameters allowing for changes in the variance of the errors. For instance, an issue of interest in macroeconomics is whether the great moderation was due to changes in the persistence parameters (the sum of the autoregressive coefficients) as suggested by the "improved policy" hypothesis or in the error variance as suggested by the "good luck" hypothesis or in both. Our tests allow to disentangle these effects, including cases with multiple breaks. Section 7 provides empirical examples related to inflation and real interest rate series. To reach the right conclusion about the number and nature of the changes, we use all tests proposed in this paper in a careful way. Obviously, the number of potential other applications abound. One could argue that it is sufficient to have tests for changes in parameters that are robust to unknown patterns of changes in variance. An example is the work of Górecki, Horváth, and Kokoszka (2018). However, their tests are based on a two-step approach; first estimating the error process assuming no coefficient breaks and subsequently testing for changes in the coefficients using this estimate. Accordingly, the tests can suffer from severe power losses as the estimated error process can be contaminated when structural changes are actually present in the coefficients. Indeed, unreported simulations show their tests to have nonmonotonic power, that is, power that decreases as the magnitude of the change in the regression parameters increases. This testing problem is easily covered via our $\sup \mathrm{LR}_{3,T}$ and $\mathrm{UD}\max \mathrm{LR}_{3,T}$ tests, which maintain good power properties. Similarly, one could be content with only testing for a change in variance allowing for unspecified changes in the regression parameters. The only tests we know that tackle this issue are based on the $\sup \mathrm{LR}_{2,T}$ and $\mathrm{UD}\max \mathrm{LR}_{2,T}$ tests that we propose.

The paper is structured as follows. Section 2 presents the models and testing problems, with the quasi-likelihood tests stated in Section 3. Section 4 discusses the assumptions needed on the regressors and errors, derives the relevant limit distributions under the various null hypotheses and proposes corrected versions of the tests that have limit distributions-free of nuisance parameters. Section 4.1 deals with the case of martingale difference errors, Section 4.2 extends the analysis to serially correlated errors, Section 4.3 covers the case with an unknown number of breaks. Section 4.4 discusses tests for an additional break in either the regression coefficients or the variance. Section 5 provides simulation results to assess the adequacy of the suggested procedures in terms of their finite sample size and power and provides some practical guidelines. Section 6 discusses methods to estimate the number of breaks in the regression coefficients and the variance. Section 7 provides empirical applications and Section 8 brief concluding remarks. An Appendix contains some technical derivations. Additional material can be found in the Online Supplemental Material in the replication file (Perron, Yamamoto, and Zhou (2020)).

2. Model and testing problems

We start with a description of the most general specification of the model considered where multiple breaks occur in both the coefficients of the conditional mean and the variance of the errors, at possibly different times. This will allow us to set up the notation used throughout the paper. The main framework of analysis can be described by the following multiple linear regression with m breaks (or $m+1$ regimes) in the conditional mean equation:

$$y_t = x_t'\beta + z_t'\delta_j + u_t, \quad t = T_{j-1}^c + 1, \ldots, T_j^c, \tag{1}$$

for $j = 1, \ldots, m+1$. In this model, y_t is the observed dependent variable at time t; both x_t $(p \times 1)$ and z_t $(q \times 1)$ are vectors of covariates and β and δ_j $(j = 1, \ldots, m+1)$ are the corresponding vectors of coefficients; u_t is the disturbance at time t. The break dates $(T_1^c, \ldots, T_m^c)$ are explicitly treated as unknown (with the convention $T_0^c = 0$ and $T_{m+1}^c = T$ used). This is a partial structural change model since the parameter vector β is not subject to shifts and is estimated using the entire sample. When $p = 0$, we obtain a pure structural change model with all coefficients subject to change. We also allow for n breaks (or $n+1$ regimes) for the variance of the errors occurring at unknown dates $(T_1^v, \ldots, T_n^v)$. Accordingly, $E(u_t) = 0$ and $E(u_t^2) = \sigma_i^2$ for $T_{i-1}^v + 1 \leq t \leq T_i^v$ $(i = 1, \ldots, n+1)$, where again we use the convention that $T_0^v = 0$ and $T_{n+1}^v = T$. We allow the breaks in the variance and in the regression coefficients to happen at different times, hence the m-vector $(T_1^c, \ldots, T_m^c)$ and the n-vector $(T_1^v, \ldots, T_n^v)$ can have all distinct elements or they can overlap partly or completely. We let K denote the total number of break dates and $\max[m, n] \leq K \leq m + n$. When the the breaks overlap completely, $m = n = K$. The multiple linear regression system (1) may be expressed in matrix form as $Y = X\beta + \bar{Z}\delta + U$, where $Y = (y_1, \ldots, y_T)'$, $X = (x_1, \ldots, x_T)'$, $U = (u_1, \ldots, u_T)'$, $\delta = (\delta_1', \ldots, \delta_{m+1}')'$, and $\bar{Z}$ diagonally partitions Z at $(T_1^c, \ldots, T_m^c)$, that is, $\bar{Z} = \operatorname{diag}(Z_1, \ldots, Z_{m+1})$ with $Z_j = (z_{T_{j-1}^c+1}, \ldots, z_{T_j^c})'$. The true value of the parameters are $\delta^0 = (\delta_1^{0'}, \ldots, \delta_{m+1}^{0'})'$ and $(T_1^{c0}, \ldots, T_m^{c0})$ and $\bar{Z}^0$ diagonally partitions Z at $(T_1^{c0}, \ldots, T_m^{c0})$. Hence, the data-generating process (DGP) is $Y = X\beta^0 + \bar{Z}^0\delta^0 + U$ with $E(UU') = \Omega^0$, where the diagonal elements of Ω^0 are σ_{i0}^2 for $T_{i-1}^{v0} + 1 \leq t \leq T_i^{v0}$ $(i = 1, \ldots, n+1)$. We also consider cases with serial correlation in the errors for which the off-diagonal elements of Ω^0 need not be 0. This is a special case of the class of models considered by Qu and Perron (2007a). Their method of estimation is quasi maximum likelihood (QML) assuming serially uncorrelated Gaussian errors. They prove consistency of the estimates of the break fractions $(\lambda_1^0, \ldots, \lambda_K^0) \equiv (T_1^0/T, \ldots, T_K^0/T)$, where T_i^0 $(i = 1, \ldots, K)$ denotes the union of the elements of $(T_1^{c0}, \ldots, T_m^{c0})$ and $(T_1^{v0}, \ldots, T_n^{v0})$. This is done under general conditions on the regressors and the errors; see Section 4. Importantly, from a practical perspective, they provide an efficient estimation algorithm, which we build upon.

The testing problems are the following: TP-1: $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = 0, n = n_a\}$; TP-2: $H_0 : \{m = m_a, n = 0\}$ versus $H_1 : \{m = m_a, n = n_a\}$; TP-3: $H_0 : \{m = 0, n = n_a\}$ versus $H_1 : \{m = m_a, n = n_a\}$; TP-4: $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = m_a, n = n_a\}$, where m_a and n_a are some positive numbers selected a priori. We shall also consider testing problems where the alternatives specify some unknown numbers of breaks, up to some maximum. These are: TP-5: $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = 0, 1 \leq n \leq N\}$; TP-6: $H_0 : \{m = m_a, n = 0\}$ versus $H_1 : \{m = m_a, 1 \leq n \leq N\}$; TP-7: $H_0 : \{m = 0, n = n_a\}$ versus $H_1 : \{1 \leq m \leq M, n = n_a\}$; TP-8: $H_0 : \{m = n = 0\}$ versus $H_1 : \{1 \leq m \leq M, 1 \leq n \leq N\}$. We shall deal with: TP-9: $\{m = m_a, n = n_a\}$ versus $H_1 : \{m = m_a + 1, n = n_a\}$; TP-10: $\{m = m_a, n = n_a\}$ versus $H_1 : \{m = m_a, n = n_a + 1\}$, where m_a and n_a nonnegative integers. These are useful to verify the adequacy of a model with some number of breaks assessing whether including one more is warranted. In Section 6, we also consider sequential testing procedures that allow estimating the number of breaks in both δ and σ^2.

3. The quasi-likelihood ratio tests

We consider the likelihood ratio (LR) tests obtained assuming normally distributed and serially uncorrelated errors, for TP-1 to TP-4. We estimate the model using the quasi-maximum likelihood estimation method (QMLE). Consider TP-1 with no change in δ ($m = q = 0$) and testing for n_a changes in σ^2. Under H_0, the log-likelihood function is

$$\log \tilde{L}_T = -(T/2)(\log 2\pi + 1) - (T/2) \log \tilde{\sigma}^2, \tag{2}$$

where $\tilde{\sigma}^2 = T^{-1} \sum_{t=1}^T (y_t - x_t'\tilde{\beta})^2$ and $\tilde{\beta} = (\sum_{t=1}^T x_t x_t')^{-1}(\sum_{t=1}^T x_t y_t)$. Under H_1, for a given partition $\{T_1^v, \dots, T_n^v\}$, the log-likelihood value is given by

$$\log \hat{L}_T(T_1^v, \dots, T_n^v) = -(T/2)(\log 2\pi + 1) - \sum_{i=1}^{n_a+1} \bigl[(T_i^v - T_{i-1}^v)/2\bigr] \log \hat{\sigma}_i^2, \tag{3}$$

where the QMLE jointly solves $\hat{\beta} = (\sum_{i=1}^{n_a+1} \sum_{t=T_{i-1}^v+1}^{T_i^v} x_t x_t'/\hat{\sigma}_i^2)^{-1}(\sum_{i=1}^{n_a+1} \sum_{t=T_{i-1}^v+1}^{T_i^v} x_t y_t/\hat{\sigma}_i^2)$ and $\hat{\sigma}_i^2 = (T_i^v - T_{i-1}^v)^{-1} \sum_{t=T_{i-1}^v+1}^{T_i^v} (y_t - x_t'\hat{\beta})^2$, for $i = 1, \dots, n_a + 1$. Hence, the Sup-LR test is

$$\begin{aligned} \sup \mathrm{LR}_{1,T}(n_a, \varepsilon | m = n = 0) &= \sup_{(\lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_{v,\varepsilon}} 2\bigl[\log \hat{L}_T(T_1^v, \dots, T_{n_a}^v) - \log \tilde{L}_T\bigr] \\ &= 2\bigl[\log \hat{L}_T(\hat{T}_1^v, \dots, \hat{T}_{n_a}^v) - \log \tilde{L}_T\bigr] \end{aligned}$$

where $(\hat{T}_1^v, \dots, \hat{T}_{n_a}^v)$ are the QMLE obtained imposing the restriction of no change in the coefficients and $\Lambda_{v,\varepsilon} = \{(\lambda_1^v, \dots, \lambda_{n_a}^v); |\lambda_{i+1}^v - \lambda_i^v| \geq \varepsilon \ (i = 1, \dots, n_a - 1), \lambda_1^v \geq \varepsilon, \lambda_{n_a}^v \leq 1 - \varepsilon\}$, with ε a truncation imposing a minimal length for each segment. For TP-2, there are m_a breaks in δ under both H_0 and H_1, so the test pertains to assess whether there are 0 or n_a breaks in variance. For a given partition $\{T_1^c, \dots, T_{m_a}^c\}$, the likelihood function under H_0 is $\log \widetilde{L}_T(T_1^c, \dots, T_{m_a}^c) = -(T/2)(\log 2\pi + 1) - (T/2) \log \tilde{\sigma}^2$, where $\tilde{\sigma}^2 = T^{-1} \sum_{t=1}^T (y_t - x_t'\tilde{\beta} - z_t'\tilde{\delta}_{t,j})^2$, $\tilde{\beta} = (X' M_{\bar{Z}} X)^{-1} X' M_{\bar{Z}} Y$ and $\widetilde{\delta}_{t,j} = (Z_j' Z_j)^{-1} Z_j(Y_j - X_j \tilde{\beta})$ for $T_{j-1}^c < t \leq T_j^c$, with $M_{\bar{Z}} = I - \bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'$, $\bar{Z} = \operatorname{diag}(Z_1, \dots, Z_{m_a+1})$, and $Z_j = (z_{T_{j-1}^c+1}, \dots, z_{T_j^c})'$, $Y_j = (y_{T_{j-1}^c+1}, \dots, y_{T_j^c})'$, $X_j = (x_{T_{j-1}^c+1}, \dots, x_{T_j^c})'$ for $T_{j-1}^c < t \leq T_j^c$ $(j = 1, \dots, m_a + 1)$. The log-likelihood value under H_1 is, for given partitions $\{T_1^c, \dots, T_{m_a}^c\}$

and $\{T_1^v, \dots, T_{n_a}^v\}$,

$$\log \hat{L}_T\big(T_1^c, \dots, T_{m_a}^c; T_1^v, \dots, T_{n_a}^v\big) = -(T/2)(\log 2\pi + 1) - \sum_{i=1}^{n_a+1} \big[\big(T_i^v - T_{i-1}^v\big)/2\big] \log \hat{\sigma}_i^2, \quad (4)$$

where the QMLE solves the following equations: $\hat{\sigma}_i^2 = (T_i^v - T_{i-1}^v)^{-1} \sum_{t=T_{i-1}^v+1}^{T_i^v} (y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j})^2$ $(i = 1, \dots, n_a + 1)$ and $\hat{\beta} = (X^{\sigma\prime} M_{\bar{Z}_\sigma} X^\sigma)^{-1} X^{\sigma\prime} M_{\bar{Z}_\sigma} Y^\sigma$, where $M_{\bar{Z}_\sigma} = I - \bar{Z}_\sigma(\bar{Z}_\sigma' \bar{Z}_\sigma)^{-1} \bar{Z}_\sigma'$ with $\bar{Z}_\sigma = \operatorname{diag}(Z_1^\sigma, \dots, Z_{m_a+1}^\sigma)$, $Z_j^\sigma = (z_{T_{j-1}^c+1}^\sigma, \dots, z_{T_j^c}^\sigma)'$, and $z_t^\sigma = (z_t/\hat{\sigma}_i)$, for $T_{i-1}^v < t \le T_i^v$ $(i = 1, \dots, n_a + 1)$. Also, $\hat{\delta}_{t,j} = (Z_j^{\sigma\prime} Z_j^\sigma)^{-1} Z_j^{\sigma\prime}(Y_j^\sigma - X_j^\sigma \hat{\beta})$ for $T_{j-1}^c < t \le T_j^c$, where $Y_j^\sigma = (y_{T_{j-1}^c+1}^\sigma, \dots, y_{T_j^c}^\sigma)'$, $X_j^\sigma = (x_{T_{j-1}^c+1}^\sigma, \dots, x_{T_j^c}^\sigma)'$ with $x_t^\sigma = (x_t/\hat{\sigma}_i)$ and $y_t^\sigma = (y_t/\hat{\sigma}_i)$. Hence,

$$\begin{aligned}
&\sup \mathrm{LR}_{2,T}(m_a, n_a, \varepsilon | n = 0, m_a) \\
&\quad = 2\Big[\sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c; \lambda_1^v, \dots, \lambda_{n_a}^v) \in \Lambda_\varepsilon} \log \hat{L}_T\big(T_1^c, \dots, T_{m_a}^c; T_1^v, \dots, T_{n_a}^v\big) \\
&\qquad - \sup_{(\lambda_1^c, \dots, \lambda_{m_a}^c) \in \Lambda_{c,\varepsilon}} \log \tilde{L}_T\big(T_1^c, \dots, T_{m_a}^c\big)\Big] \\
&\quad = 2\big[\log \hat{L}_T\big(\tilde{T}_1^c, \dots, \tilde{T}_{m_a}^c; \tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v\big) - \log \tilde{L}_T\big(\hat{T}_1^c, \dots, \hat{T}_{m_a}^c\big)\big],
\end{aligned}$$

where $\Lambda_{c,\varepsilon} = \{(\lambda_1^c, \dots, \lambda_m^c); |\lambda_{j+1}^c - \lambda_j^c| \ge \varepsilon \ (j = 1, \dots, m_a - 1), \lambda_1^c \ge \varepsilon, \lambda_{m_a}^c \le 1 - \varepsilon\}$ and

$$\begin{aligned}
\Lambda_\varepsilon = \big\{&\big(\lambda_1^c, \dots, \lambda_m^c, \lambda_1^v, \dots, \lambda_n^v\big); \text{for } (\lambda_1, \dots, \lambda_K) = \big(\lambda_1^c, \dots, \lambda_m^c\big) \cup \big(\lambda_1^v, \dots, \lambda_n^v\big), \\
&|\lambda_{j+1} - \lambda_j| \ge \varepsilon \ (j = 1, \dots, K - 1), \lambda_1 \ge \varepsilon, \lambda_K \le 1 - \varepsilon\big\}. \quad (5)
\end{aligned}$$

Note that we denote the estimates of the break dates in coefficients and variance by a "$\sim$" when these are obtained jointly, and by a "$\wedge$" when obtained separately.

The set Λ_ε which defines the possible values of the break fractions in δ $(\lambda_1^c, \dots, \lambda_m^c)$ and in σ^2 $(\lambda_1^v, \dots, \lambda_m^v)$ allows them to have some (or all) common elements or be completely different. What is important is that each break fraction be separated by some $\varepsilon > 0$. This does complicate inference since many cases need to be considered. To illustrate, consider $m_a = n_a = 1$. We can have $K = 1$, a one break model with both δ and σ^2 changing at the same date. On the other hand, if $K = 2$, the break date for the change in δ is different from that for the change in σ^2. This leads to two additional possible cases: (a) $\lambda_1^c \le \lambda_1^v - \varepsilon$ (the break in δ is before that in σ^2), (b) $\lambda_1^c \ge \lambda_1^v + \varepsilon$ (the break in δ is after that in σ^2). The maximized likelihood function for these two cases can be evaluated using the algorithm of Qu and Perron (2007a) since it permits imposing restrictions. For example, if $\lambda_1^c \le \lambda_1^v - \varepsilon$, we have a two break model and the restrictions are that the error variances in the first and second regimes are identical, and the coefficients are the same in the second and third regimes. Hence, for the case $m_a = n_a = 1$, there are three maximized likelihood values to construct and the test corresponds to the maximal value over these three cases. When m_a or n_a are greater than one, more cases need to be considered, but the principle is the same.

For TP-3, H_0 specifies n_a breaks in σ^2 and none in δ. For a partition $\{T_1^v, \dots, T_{n_a}^v\}$, the likelihood function is $\log \widetilde{L}_T(T_1^v, \dots, T_{n_a}^v) = -(T/2)(\log 2\pi + 1) - \sum_{i=1}^{n_a+1}[(T_i^v - T_{i-1}^v)/2]\log \tilde{\sigma}_i^2$, where $\tilde{\sigma}_i^2 = (T_i^v - T_{i-1}^v)^{-1}\sum_{t=T_{i-1}^v+1}^{T_i^v}(y_t - x_t'\tilde{\beta} - z_t'\tilde{\delta})^2$ for $i = 1, \dots, n_a + 1$, with $(\tilde{\beta}', \tilde{\delta}')' = (W^{\sigma\prime}W^{\sigma})^{-1}W^{\sigma\prime}Y^{\sigma}$, $W^{\sigma} = (w_1^{\sigma}, \dots, w_T^{\sigma})'$ and $w_t^{\sigma} = (x_t^{\sigma\prime}, z_t^{\sigma\prime})'$. Under H_1, there are m_a breaks in δ and n_a breaks in σ^2 and the likelihood function is (4). The sup-LR test is

$$
\begin{aligned}
&\sup \mathrm{LR}_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a)\\
&= 2\Big[\sup_{(\lambda_1^c,\dots,\lambda_{m_a}^c;\lambda_1^v,\dots,\lambda_{n_a}^v)\in\Lambda_\varepsilon} \log \hat{L}_T\big(T_1^c, \dots, T_{m_a}^c; T_1^v, \dots, T_{n_a}^v\big)\\
&\quad - \sup_{(\lambda_1^v,\dots,\lambda_{n_a}^v)\in\Lambda_{v,\varepsilon}} \log \tilde{L}_T\big(T_1^v, \dots, T_{n_a}^v\big)\Big]\\
&= 2\big[\log \hat{L}_T\big(\tilde{T}_1^c, \dots, \tilde{T}_{m_a}^c; \tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v\big) - \log \tilde{L}_T\big(\hat{T}_1^v, \dots, \hat{T}_{n_a}^v\big)\big].
\end{aligned}
$$

For TP-4, under H_0 we have no break and the log-likelihood function is (2). H_1 specifies m_a breaks in δ and n_a breaks in σ^2 and the log likelihood is (4). Hence, the Sup-LR test is

$$
\begin{aligned}
&\sup \mathrm{LR}_{4,T}(m_a, n_a, \varepsilon | n = m = 0)\\
&= 2\Big[\sup_{(\lambda_1^c,\dots,\lambda_{m_a}^c;\lambda_1^v,\dots,\lambda_{n_a}^v)\in\Lambda_\varepsilon} \log \hat{L}_T\big(T_1^c, \dots, T_{m_a}^c; T_1^v, \dots, T_{n_a}^v\big) - \log \tilde{L}_T\Big]\\
&= 2\big[\log \hat{L}_T\big(\tilde{T}_1^c, \dots, \tilde{T}_{m_a}^c; \tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v\big) - \log \tilde{L}_T\big]. \qquad (6)
\end{aligned}
$$

4. The limiting distributions of the tests

The limit distribution of the tests for martingale difference errors is presented in Section 4.1 with extensions to serially correlated errors in Section 4.2. Section 4.3 deals with double maximum tests and Section 4.4 with tests for an additional break; "$\to_p$" denotes convergence in probability, "$\Rightarrow$" weak convergence under the Skorohod topology and $\|\cdot\|$ is the Euclidean norm.

4.1 *The case with martingale difference errors*

When σ^2 is constant under H_0 but allowed to change under H_1 (TP-1, 2, 4), we specify:

• Assumption A1. *The errors $\{u_t\}$ form an array of martingale differences relative to $\mathcal{F}_t = \sigma$-field$\{\dots, z_{t-1}, z_t, \dots, x_{t-1}, x_t, \dots, u_{t-2}, u_{t-1}\}$, $E(u_t^2) = \sigma_0^2$ for all t and $T^{-1/2} \times \sum_{t=1}^{[Ts]}(u_t^2/\sigma_0^2 - 1) \Rightarrow \psi W(s)$, where $W(s)$ is a Wiener process and $\psi = \lim_{T\to\infty} \mathrm{var}(T^{-1/2} \times \sum_{t=1}^{T}(u_t^2/\sigma_0^2 - 1))$.*

Assumption A1 rules out instability in the error process and states that a basic functional central limit theorem holds for the partial sums of the squared errors. When changes in the coefficients are tested (TP-3 and TP-4), we assume, with $w_t = (x_t', z_t')'$:

• Assumption A2. *The errors* $\{u_t\}$ *form an array of martingale differences relative to* $\mathcal{F}_t = \sigma$-field$\{\dots, z_{t-1}, z_t, \dots, x_{t-1}, x_t, \dots, u_{t-2}, u_{t-1}\}$, $T^{-1}\sum_{t=1}^{[Ts]} w_t w_t' \to_p sQ$,*uniformly in* $s \in [0, 1]$, *with* Q *some positive definite matrix and* $T^{-1/2}\sum_{t=1}^{[Ts]} z_t u_t \Rightarrow \sigma_0 Q^{1/2} W_q(s)$, *where* $W_q(s)$ *is a* q*-vector of independent Wiener processes independent of* $W(s)$.

The first part of Assumption A2 rules out trending regressors and requires the limit moment matrix of the regressors be homogeneous throughout the sample. Hence, we avoid changes in the marginal distribution of the regressors when the coefficients do not change (e.g., Hansen (2000), Cavaliere and Georgiev (2019)). This follows from our basic premise that regimes are defined by changes in some coefficients. The second part of Assumption A2 assumes no serial correlation in the errors u_t but this will be relaxed later. Since some testing problems imply a nonzero number of breaks under H_0, that is, in TP-2 and TP-3, we need the following conditions to ensure that the estimates of the break fractions are consistent at a fast enough rate so that they do not affect the distributions of the parameters asymptotically. This problem was analyzed in Qu and Perron (2007a) and we simply use the same set of assumptions:

• Assumption A3. *The conditions stated in Assumptions* A1–A9 *of Qu and Perron* (2007a) *are assumed to hold with the segments defined for* T_i^0 $(i = 1, \dots, K)$. *However,* A6 *is replaced by (for* $j = 1, \dots, m$ *and* $i = 1, \dots, n$*):* $\delta_{j+1}^0 - \delta_j^0 = v_T^\delta \delta_j^*$ *and* $\sigma_{i+1,0} - \sigma_{i,0} = v_T \sigma_{i,0}^*$, *where* $(\delta_j^*, \sigma_{i,0}^*) \neq 0$ *and are independent of* T. *Moreover,* v_T^δ *is either a positive number independent of* T *or a sequence of positive numbers satisfying* $v_T^\delta \to 0$ *and* $T^{1/2} v_T^\delta / (\log T)^2 \to \infty$, *while* v_T *is a sequence of positive numbers satisfying* $v_T \to 0$ *and* $T^{1/2} v_T / (\log T)^2 \to \infty$.

The main difference is that we require the changes in the variance of the errors to decrease to 0 at a slow enough rate as T increases, while the changes in the coefficients can be fixed or decreasing. Both cases ensure that the estimates of the break fractions are consistent and that the limit distributions of the parameter estimates are the same as when the true break dates are known. The requirement that the change in variance must decrease as T increases is to ensure that Assumption A2 holds when changes in variance are permitted under the null hypothesis, in particular if lagged dependent variables are present. Otherwise, the limit distribution of the test for TP-3 is not invariant to nuisance parameters. This is not constraining in practice since the rate of decrease can be as slow as desired. We will show via simulations that the exact size of the test is close to the nominal level whether the changes in variance are small or large. To see why this is needed to ensure that Assumption A2 is satisfied, let $z_t u_t^\sigma = z_t u_t / \sigma_{i0}$. Then

$$T^{-1/2}\sum_{t=1}^{[Ts]} z_t u_t = T^{-1/2}\sigma_0 \sum_{t=1}^{[Ts]} z_t u_t^\sigma + \sum_{i=1}^{n_a+1}\left(\frac{\sigma_{i0} - \sigma_0}{\sigma_{i0}}\right)\left(T^{-1/2}\sum_{t=T_{i-1}^{v0}+1}^{T_i^{v0}} z_t u_t\right) \Rightarrow \sigma_0 Q^{1/2} W_q(s),$$

where $\sigma_0 = \sigma_{10}$ without loss of generality. The result follows since $[(\sigma_{i0} - \sigma_0)/\sigma_{i0}] = O_p(v_T)$, $v_T \to 0$ and $T^{-1/2}\sum_{t=T_{i-1}^{v0}+1}^{T_i^{v0}} z_t u_t = O_p(1)$. The same applies to the requirement that $T^{-1}\sum_{t=1}^{[Ts]} w_t w_t' \to_p sQ$ uniformly in s. To see that this holds when lagged dependent variables are present, consider a simple AR(1) model $y_t = \beta y_{t-1} + u_t$ in which σ^2 has n breaks and $|\beta| < 1$. Using the variance adjusted series $y_t^\sigma = \beta y_{t-1}^\sigma + u_t^\sigma$ where $u_t^\sigma = u_t/\sigma_{i0}$, we have

$$T^{-1}\sum_{t=1}^{[Ts]} z_t z_t' = T^{-1}\sum_{t=1}^{[Ts]} y_{t-1}^2 = T^{-1}\sigma_0^2\sum_{t=1}^{[Ts]} y_{t-1}^{\sigma 2} + O_p(\upsilon_T) \xrightarrow{p} sQ, \tag{7}$$

where $Q = \sigma_0^2/(1-\beta^2)$ (see Supplement A). Why υ_T^δ can remain fixed when δ changes is because such breaks do not affect the moments of the errors, and when lagged dependent variables are present changes in δ imply changes in the marginal distribution of the regressors (e.g., the lagged dependent variables) occurring at the same times, which is allowed. The limiting distributions of the LR tests under H_0, are stated in the following theorem.

Theorem 1. *Under the relevant null H_0, we have, as $T \to \infty$:* (a) *For TP-*1*, under Assumption* A1:

$$\sup \mathrm{LR}_{1,T}(n_a, \varepsilon | m = n = 0) \Rightarrow \sup_{(\lambda_1^v,\ldots,\lambda_{n_a}^v)\in\Lambda_{v,\varepsilon}} \frac{\psi}{2}\sum_{i=1}^{n_a}\frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v)\right)^2}{\lambda_{i+1}^v\lambda_i^v(\lambda_{i+1}^v - \lambda_i^v)}.$$

(b) *For TP-*2*, under Assumptions* A1 *and* A3*,*

$$\begin{aligned}\sup \mathrm{LR}_{2,T}(m_a, n_a, \varepsilon | n = 0, m_a) &\Rightarrow \sup_{(\lambda_1^v,\ldots,\lambda_{n_a}^v)\in\Lambda_{v,\varepsilon}^c} \frac{\psi}{2}\sum_{i=1}^{n_a}\frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v)\right)^2}{\lambda_{i+1}^v\lambda_i^v(\lambda_{i+1}^v - \lambda_i^v)}\\ &\leq \sup_{(\lambda_1^v,\ldots,\lambda_{n_a}^v)\in\Lambda_{v,\varepsilon}} \frac{\psi}{2}\sum_{i=1}^{n_a}\frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v)\right)^2}{\lambda_{i+1}^v\lambda_i^v(\lambda_{i+1}^v - \lambda_i^v)},\end{aligned}$$

where $\Lambda_{v,\varepsilon}^c = \{(\lambda_1^v, \ldots, \lambda_n^v);$ *for* $(\lambda_1, \ldots, \lambda_K) = (\lambda_1^{c0}, \ldots, \lambda_m^{c0}) \cup (\lambda_1^v, \ldots, \lambda_n^v), |\lambda_{j+1} - \lambda_j| \geq \varepsilon\ (j = 1, \ldots, K-1), \lambda_1 \geq \varepsilon, \lambda_K \leq 1-\varepsilon\}$. (c) *For TP-*3*, under Assumptions* A2 *and* A3:

$$\begin{aligned}\sup \mathrm{LR}_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a) &\Rightarrow \sup_{(\lambda_1^c,\ldots,\lambda_{m_a}^c)\in\Lambda_{c,\varepsilon}^v} \sum_{j=1}^{m_a}\frac{\left\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\right\|^2}{\lambda_{j+1}^c\lambda_j^c(\lambda_{j+1}^c - \lambda_j^c)}\\ &\leq \sup_{(\lambda_1^c,\ldots,\lambda_{m_a}^c)\in\Lambda_{c,\varepsilon}} \sum_{j=1}^{m_a}\frac{\left\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\right\|^2}{\lambda_{j+1}^c\lambda_j^c(\lambda_{j+1}^c - \lambda_j^c)},\end{aligned}$$

where $\Lambda_{c,\varepsilon}^v = \{(\lambda_1^c, \ldots, \lambda_m^c);$ *for* $(\lambda_1, \ldots, \lambda_K) = (\lambda_1^c, \ldots, \lambda_m^c) \cup (\lambda_1^{v0}, \ldots, \lambda_n^{v0}), |\lambda_{j+1} - \lambda_j| \geq \varepsilon\ (j = 1, \ldots, K-1), \lambda_1 \geq \varepsilon, \lambda_K \leq 1-\varepsilon\}$. (d) *For TP-*4*, under Assumptions* A1 *and* A2:

$$\begin{aligned}\sup \mathrm{LR}_{4,T}(m_a, n_a, \varepsilon | n = m = 0) \Rightarrow \sup_{(\lambda_1^c,\ldots,\lambda_{m_a}^c;\lambda_1^v,\ldots,\lambda_{n_a}^v)\in\Lambda_\varepsilon} &\left[\sum_{j=1}^{m_a}\frac{\left\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\right\|^2}{\lambda_{j+1}^c\lambda_j^c(\lambda_{j+1}^c - \lambda_j^c)}\right.\\ &\left. + \frac{\psi}{2}\sum_{i=1}^{n_a}\frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v)\right)^2}{\lambda_{i+1}^v\lambda_i^v(\lambda_{i+1}^v - \lambda_i^v)}\right]\end{aligned}$$

$$\leq \sup_{(\lambda_1^c,\ldots,\lambda_{m_a}^c;\lambda_1^v,\ldots,\lambda_{n_a}^v)\in\Lambda_{cv,\varepsilon}} \left[\sum_{j=1}^{m_a} \frac{\left\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\right\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} + \frac{\psi}{2} \sum_{i=1}^{n_a} \frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v)\right)^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)}\right],$$

where $\Lambda_{cv,\varepsilon} = \{(\lambda_1^c, \ldots, \lambda_m^c; \lambda_1^v, \ldots, \lambda_{n_a}^v); |\lambda_{j+1}^c - \lambda_j^c| \geq \varepsilon\ (j = 1, \ldots, m_a - 1), \lambda_1^c \geq \varepsilon, \lambda_{m_a}^c \leq 1 - \varepsilon, |\lambda_{i+1}^v - \lambda_i^v| \geq \varepsilon\ (i = 1, \ldots, n_a - 1), \lambda_1^v \geq \varepsilon, \lambda_{n_a}^v \leq 1 - \varepsilon\}$.

Except for TP-1, the limit distributions depend on the interval between the break fractions for δ and σ^2 when they do not coincide. This imposes restrictions on the parameter space of the break fractions. Hence, the critical values are smaller than what is obtained from the standard limit distribution in Bai and Perron (1998). Although the computation of such limit distributions might be feasible, it is beyond the scope of this study. The results, however, show that these distributions are bounded by limit random variables which can easily be simulated. This follows since $\Lambda_{v,\varepsilon}^c \subseteq \Lambda_{v,\varepsilon}$, $\Lambda_{c,\varepsilon}^v \subseteq \Lambda_{c,\varepsilon}$, and $\Lambda_\varepsilon \subseteq \Lambda_{cv,\varepsilon}$. Hence, a conservative testing procedure is possible. As we shall see, the test is barely conservative if the trimming parameter ε is small, though as ε gets large (e.g., 0.20) the test will be somewhat undersized. The proof of this theorem is given in the Appendix. For TP-3, the bound is the same as the limit distribution in Bai and Perron (1998, 2003b) and the critical values they provided can be used. For TP-1 and TP-2, the same limit distribution (for a one parameter change) applies except for the scaling factor $(\psi/2)$. This quantity can nevertheless still be consistently estimated. Consider the class of estimates:

$$\hat{\psi} = T^{-1} \sum_{j=-(T-1)}^{T-1} \omega(j, b_T) \sum_{t=|j|+1}^{T} \hat{\eta}_t \hat{\eta}_{t-j}, \tag{8}$$

where $\hat{\eta}_t = (\hat{u}_t^2/\hat{\sigma}^2) - 1$ and $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^T \hat{u}_t^2$ with $\hat{u}_t$ the estimated residuals. Here, $\omega(j, b_T)$ is a weight function and b_T some selected bandwidth. The estimate $\hat{\psi}$ will be consistent under some conditions on the choice of $\omega(j, b_T)$ and the rate of increase of b_T as a function of T. Following Kejriwal (2009) (see also Kejriwal and Perron (2010)), we use the residuals under H_0 to construct the sample autocovariances of η_t but the residuals under H_1 to select the bandwidth parameter b_T; see Supplement B for details. In our simulations and empirical applications, we use the quadratic spectral kernel and to select b_T we use the method of Andrews (1991) with an AR(1) approximation. If the errors are i.i.d., $\psi = \mu_4/\sigma^4 - 1$, which can be consistently estimated using $\hat{\psi} = \hat{\mu}_4/\hat{\sigma}^4 - 1$, where $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^T \hat{u}_t^2$ and $\hat{\mu}_4 = T^{-1}\sum_{t=1}^T \hat{u}_t^4$ with $\hat{u}_t$ the residuals under the null or alternative hypotheses. Also, if the errors are normal as in Qu and Perron (2007a), $\psi = 2$ so that no adjustment is necessary. We shall only consider a correction involving $\hat{\psi}$ as defined by (8) for all cases; Supplement C shows that there is no loss in power in doing so and that the size remains adequate. The following corrected statistics then have nuisance parameter-free limit distributions:

$$
\begin{aligned}
\sup \mathrm{LR}^*_{1,T} = (2/\hat{\psi}) \sup \mathrm{LR}_{1,T} &\Rightarrow \sup_{(\lambda^v_1,\dots,\lambda^v_{na})\in\Lambda_{v,\varepsilon}} \sum_{i=1}^{n_a} \frac{\left(\lambda^v_i W\left(\lambda^v_{i+1}\right) - \lambda^v_{i+1} W\left(\lambda^v_i\right)\right)^2}{\lambda^v_{i+1}\lambda^v_i\left(\lambda^v_{i+1} - \lambda^v_i\right)}, \\
\sup \mathrm{LR}^*_{2,T} = (2/\hat{\psi}) \sup \mathrm{LR}_{2,T} &\Rightarrow \sup_{(\lambda^v_1,\dots,\lambda^v_{na})\in\Lambda^c_{v,\varepsilon}} \sum_{i=1}^{n_a} \frac{\left(\lambda^v_i W\left(\lambda^v_{i+1}\right) - \lambda^v_{i+1} W\left(\lambda^v_i\right)\right)^2}{\lambda^v_{i+1}\lambda^v_i\left(\lambda^v_{i+1} - \lambda^v_i\right)} \\
&\le \sup_{(\lambda^v_1,\dots,\lambda^v_{na})\in\Lambda_{v,\varepsilon}} \sum_{i=1}^{n_a} \frac{\left(\lambda^v_i W\left(\lambda^v_{i+1}\right) - \lambda^v_{i+1} W\left(\lambda^v_i\right)\right)^2}{\lambda^v_{i+1}\lambda^v_i\left(\lambda^v_{i+1} - \lambda^v_i\right)}.
\end{aligned}
\tag{9}
$$

For TP-4, it is possible to obtain a transformation with a limit distribution-free of nuisance parameters but the procedure is more involved. It is given by

$$
\sup \mathrm{LR}^*_{4,T} = \sup \mathrm{LR}_{4,T} - \left[(\hat{\psi} - 2)/\hat{\psi}\right] \mathrm{LR}_v, \tag{10}
$$

where LR_v is the LR test for 0 versus n_a breaks in variance evaluated at $\{\tilde{T}^v_1, \dots, \tilde{T}^v_{n_a}\}$ obtained by maximizing the likelihood function jointly allowing for m_a breaks in δ, that is,

$$
\mathrm{LR}_v = 2\left[\log \hat{L}_T\left(\tilde{T}^v_1, \dots, \tilde{T}^v_{n_a}\right) - \log \tilde{L}_T\right], \tag{11}
$$

where $\log \hat{L}_T(\cdot)$ and $\log \tilde{L}_T$ are defined by (3) and (2), respectively. Note that LR_v is not equivalent to $\mathrm{LR}_{1,T}(n_a, \varepsilon | m = n = 0)$ which is based on the estimates of the break dates for the changes in variance assuming no break in coefficients. Since $\{\tilde{T}^v_1/T, \dots, \tilde{T}^v_{n_a}/T\}$ are consistent estimates of the break fractions whether $m_a = 0$ or not, we have

$$
\mathrm{LR}_v \Rightarrow (\psi/2) \sup_{(\lambda^v_1,\dots,\lambda^v_{na})\in\Lambda_{\varepsilon}} \sum_{i=1}^{n_a} \frac{\left(\lambda^v_i W\left(\lambda^v_{i+1}\right) - \lambda^v_{i+1} W\left(\lambda^v_i\right)\right)^2}{\lambda^v_{i+1}\lambda^v_i\left(\lambda^v_{i+1} - \lambda^v_i\right)}
$$

and hence,

$$
\begin{aligned}
\sup \mathrm{LR}^*_{4,T} &\Rightarrow \sup_{(\lambda^c_1,\dots,\lambda^c_{ma};\lambda^v_1,\dots,\lambda^v_{na})\in\Lambda_{\varepsilon}} \left[\sum_{j=1}^{m_a} \frac{\left\| \lambda^c_j W_q\left(\lambda^c_{j+1}\right) - \lambda^c_{j+1} W_q\left(\lambda^c_j\right)\right\|^2}{\lambda^c_{j+1}\lambda^c_j\left(\lambda^c_{j+1} - \lambda^c_j\right)} \right. \\
&\qquad \left. + \sum_{i=1}^{n_a} \frac{\left(\lambda^v_i W\left(\lambda^v_{i+1}\right) - \lambda^v_{i+1} W\left(\lambda^v_i\right)\right)^2}{\lambda^v_{i+1}\lambda^v_i\left(\lambda^v_{i+1} - \lambda^v_i\right)}\right] \\
&\le \sup_{(\lambda^c_1,\dots,\lambda^c_{ma};\lambda^v_1,\dots,\lambda^v_{na})\in\Lambda_{cv,\varepsilon}} \left[\sum_{j=1}^{m_a} \frac{\left\| \lambda^c_j W_q\left(\lambda^c_{j+1}\right) - \lambda^c_{j+1} W_q\left(\lambda^c_j\right)\right\|^2}{\lambda^c_{j+1}\lambda^c_j\left(\lambda^c_{j+1} - \lambda^c_j\right)} \right. \\
&\qquad \left. + \sum_{i=1}^{n_a} \frac{\left(\lambda^v_i W\left(\lambda^v_{i+1}\right) - \lambda^v_{i+1} W\left(\lambda^v_i\right)\right)^2}{\lambda^v_{i+1}\lambda^v_i\left(\lambda^v_{i+1} - \lambda^v_i\right)}\right].
\end{aligned}
\tag{12}
$$

The limit distribution (12) is new and we obtain the asymptotic critical values via simulations. The Wiener processes $W_q(\lambda)$ and $W(\lambda)$ are approximated by the partial sums $T^{-1/2}\sum_{t=1}^{[T\lambda]} e_t$ and $T^{-1/2}\sum_{t=1}^{[T\lambda]} \epsilon_t$ with $e_t \sim$ i.i.d. $N(0, I_q)$ and $\epsilon_t \sim$ i.i.d. $N(0, 1)$ which

TABLE 1. Asympotic critical values of the upper bound of the $\sup \text{LR}^*_{4,T}$ test.

		$\varepsilon = 0.10$				$\varepsilon = 0.15$				$\varepsilon = 0.20$			$\varepsilon = 0.25$	UD max LR^*_4		
		$n_a = 1$		$n_a = 2$		$n_a = 1$		$n_a = 2$		$n_a = 1$		$n_a = 2$	$n_a = 1$	$M = N = 2$		
q	α	$m_a = 1$	$m_a = 2$	$m_a = 1$	$m_a = 2$	$m_a = 1$	$m_a = 2$	$m_a = 1$	$m_a = 2$	$m_a = 1$	$m_a = 2$	$m_a = 1$	$m_a = 1$	$\varepsilon = 0.10$	$\varepsilon = 0.15$	$\varepsilon = 0.20$
1	0.90	6.59	6.34	6.32	6.20	6.21	5.75	5.72	5.46	5.83	5.19	5.18	5.48	7.18	6.61	6.15
	0.95	7.63	7.12	7.10	6.83	7.18	6.49	6.46	6.13	6.79	5.93	5.89	6.43	8.03	7.51	7.05
	0.975	8.54	7.78	7.75	7.44	8.12	7.17	7.23	6.71	7.70	6.56	6.70	7.42	8.81	8.32	7.87
	0.99	9.79	8.73	8.70	8.17	9.24	7.98	8.00	7.45	8.83	7.42	7.52	8.56	10.00	9.42	8.95
2	0.90	7.88	7.96	7.18	7.41	7.45	7.31	6.54	6.66	7.10	6.72	6.01	6.70	8.47	7.93	7.39
	0.95	8.87	8.78	7.94	8.03	8.45	8.12	7.36	7.33	8.12	7.52	6.77	7.72	9.37	8.88	8.42
	0.975	9.85	9.52	8.69	8.69	9.45	8.91	8.02	7.88	9.08	8.34	7.50	8.69	10.32	9.77	9.40
	0.99	11.12	10.55	9.52	9.52	10.73	9.90	8.93	8.73	10.27	9.31	8.33	9.94	11.47	10.96	10.54
3	0.90	8.98	9.34	7.93	8.44	8.53	8.63	7.30	7.63	8.09	7.94	6.70	7.67	9.73	9.09	8.55
	0.95	10.06	10.23	8.72	9.11	9.52	9.51	8.07	8.31	9.11	8.77	7.50	8.75	10.66	10.08	9.48
	0.975	11.08	10.98	9.43	9.75	10.61	10.30	8.80	8.98	10.18	9.59	8.25	9.73	11.48	10.93	10.41
	0.99	12.43	12.01	10.33	10.53	11.87	11.30	9.67	9.80	11.50	10.50	9.09	10.89	12.66	12.19	11.64
4	0.90	9.96	10.60	8.54	9.32	9.51	9.90	7.87	8.56	9.09	9.17	7.31	8.66	10.88	10.24	9.64
	0.95	11.10	11.51	9.38	10.05	10.54	10.83	8.73	9.30	10.14	10.01	8.14	9.73	11.85	11.19	10.66
	0.975	12.17	12.30	10.13	10.72	11.61	11.62	9.47	9.98	11.17	10.89	8.91	10.87	12.81	12.20	11.53
	0.99	13.50	13.36	11.07	11.59	13.08	12.62	10.42	10.73	12.67	11.90	9.76	12.33	13.99	13.39	12.84
5	0.90	10.94	11.81	9.19	10.21	10.45	11.03	8.53	9.41	9.99	10.36	7.94	9.56	12.07	11.33	10.70
	0.95	12.14	12.76	10.00	10.99	11.66	12.01	9.33	10.13	11.20	11.33	8.75	10.73	13.06	12.38	11.84
	0.975	13.22	13.68	10.74	11.63	12.72	12.89	10.09	10.82	12.28	12.22	9.54	11.93	13.99	13.38	12.86
	0.99	14.47	14.66	11.77	12.50	14.06	14.13	11.15	11.67	13.56	13.29	10.52	13.23	15.16	14.50	13.95

Note: The entries are quantiles x such that $P((n_a + m_a)^{-1} \sup \text{LR}^*_4 \le x) \ge \alpha$.

are mutually independent. The number of replications is 10,000 and $T = 1000$. For each replication, a sum of the supremum of $\sum_{j=1}^{m_a}(\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2)/(\lambda_{j+1}^c \times \lambda_j^c(\lambda_{j+1}^c - \lambda_j^c))$ with respect to $(\lambda_1^c, \ldots, \lambda_{m_a}^c)$ and that of $\sum_{i=1}^{n_a}((\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2)/(\lambda_{i+1}^v \lambda_i^v(\lambda_{i+1}^v - \lambda_i^v))$ with respect to $(\lambda_1^v, \ldots, \lambda_{n_a}^v)$ is obtained via a dynamic programming algorithm. The critical values for tests of size 1%, 2.5%, 5%, and 10% are presented in Table 1 for q between 1 and 5 and $\varepsilon = 0.1, 0.15, 0.20$, and 0.25. For $\varepsilon = 0.1, 0.15, 0.2$, $m_a = 1, 2$, and $n_a = 1, 2$. For $\varepsilon = 0.25$, $m_a = 1$, and $n_a = 1$ given that $\varepsilon = 0.25$ imposes a maximal number of 2 breaks.

4.2 *Extensions to serially correlated errors*

We now consider the case with serially correlated errors. For TP-1 and TP-2, the results are the same and the $\sup \text{LR}^*_{1,T}$ and $\sup \text{LR}^*_{2,T}$ statistics are asymptotically invariant to nonnormal errors, serial correlation and conditional heteroskedasticity so that the limit distribution (9) still applies. For TP-3 and TP-4, things are more complex. For TP-3, the LR type test for changes in δ depends on nuisance parameters. We suggest the following robust Wald-type statistic: $\sup_{(\lambda_1^c, \ldots, \lambda_{m_a}^c) \in \Lambda_\varepsilon} W_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a)$, where

$$W_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a) = T\hat{\delta}' R' \big(R\hat{V}(\hat{\delta})R'\big)^{-1} R\hat{\delta} \tag{13}$$

with $\hat{\delta} = (\hat{\delta}_1', \ldots, \hat{\delta}_{m_a+1}')'$ the QMLE of δ under a given partition of the sample, R is the conventional matrix such that $(R\delta)' = (\delta_1' - \delta_2', \ldots, \delta_{m_a}' - \delta_{m_a+1}')$ and $\hat{V}(\hat{\delta})$ is an estimate of the covariance matrix of $\hat{\delta}$ robust to serial correlation and heteroskedasticity, that is, a consistent estimate of $V(\hat{\delta}) = \text{plim}_{T\to\infty} T(\bar{Z}^{*\prime}_{\sigma}\bar{Z}^{*}_{\sigma})^{-1}\Omega_{\bar{Z}^{*}_{\sigma}}(\bar{Z}^{*\prime}_{\sigma}\bar{Z}^{*}_{\sigma})^{-1}$, where $\bar{Z}^{*}_{\sigma} = M_{X_\sigma}\bar{Z}_{\sigma}$, $\Omega_{\bar{Z}^{*}_{\sigma}} = E(\bar{Z}^{*\prime}_{\sigma}U^{*}_{b}U^{*\prime}_{b}\bar{Z}^{*}_{\sigma})$, $U^{*}_{b} = M_{X_\sigma}U_{\sigma}$, $M_{X_\sigma} = I_T - X_{\sigma}(X'_{\sigma}X_{\sigma})^{-1}X'_{\sigma}$, with $\bar{Z}_{\sigma} = \text{diag}(Z^{\sigma}_{1}, \dots, Z^{\sigma}_{m_a+1})$, $Z^{\sigma}_{j} = (z^{\sigma}_{T^c_{j-1}+1}, \dots, z^{\sigma}_{T^c_j})'$, $U_{\sigma} = (u^{\sigma}_{1}, \dots, u^{\sigma}_{T})'$, $z^{\sigma}_{t} = (z_t/\hat{\sigma}_i)$ and $u^{\sigma}_{t} = (u_t/\hat{\sigma}_i)$, for $T^{v0}_{i-1} < t \leq T^{v0}_{i}$ $(i = 1, \dots, n_a + 1)$. In practice, the computation of this test can be very involved. Following Bai and Perron (1998), we suggest first to use the dynamic programming algorithm to get the break points corresponding to the global maximizers of the likelihood function defined by (4), then plug the estimates into (13) to construct the test. This will not affect the consistency of the test since the break fractions are consistently estimated.

For TP-4, potential serial correlations in both u_t and η_t must be accounted for. This can easily be achieved since $\sup \text{LR}_{4,T}$ is asymptotically equivalent to $\sup \text{LR}^{*}_{4,T} = \sup \text{LR}_{3,T} + \text{LR}_{v}$. Because of the block diagonality of the information matrix, corrections can be applied to each component separately. The first term is constructed as discussed above, namely $W_{3,T}$ defined by (13), except that one can use z_t instead of z^{σ}_{t} since H_0 specifies no break in variance. The second term LR_{v} is as defined by (11) with $\hat{\psi}$ constructed by (8).

4.3 *Double maximum tests*

The tests discussed above need prior information about H_1, that is, the number of breaks in δ and in σ^2, which may be unknown. Hence the need for TP-5 to TP-8. Bai and Perron (1998) proposed double maximum tests to solve this problem with only breaks in δ. They are tests of no break against an unknown number of breaks given some upper bound. We shall only consider their UD max test. The double maximum tests can play a significant role in testing for structural changes and it is arguably the most useful test to apply when trying to determine if structural changes are present. While tests for one break are consistent against multiple changes, their power in finite samples can sometimes be poor. There are types of multiple structural changes that are difficult to detect with a test for a single change (e.g., two breaks with the first and third regimes the same). Also, tests for a particular number of changes may have nonmonotonic power when the number of changes is greater than specified. Furthermore, the simulations of Bai and Perron (2006) show, in the context of testing for changes in the regression coefficients, that the power of the double maximum tests is almost as high as the best power achievable using the test specified with the correct number of breaks. All these elements strongly point to their usefulness. For each testing problem, the tests and their limit distributions are presented in the following theorem.

THEOREM 2. *Under the relevant H_0, we have, as $T \to \infty$:* (a) *For TP-5, under Assumption A1:*

$$\begin{aligned} \text{UD}\max \text{LR}_{1,T} &= \max_{1\leq n_a\leq N} n_a^{-1} \sup \text{LR}^{*}_{1,T}(n_a, \varepsilon | m = n = 0) \\ &\Rightarrow \max_{1\leq n_a\leq N} n_a^{-1} \sup_{(\lambda^{v}_{1},\dots,\lambda^{v}_{n_a})\in\Lambda_{v,\varepsilon}} \sum_{i=1}^{n_a} \frac{\left(\lambda^{v}_{i}W(\lambda^{v}_{i+1}) - \lambda^{v}_{i+1}W(\lambda^{v}_{i})\right)^2}{\lambda^{v}_{i+1}\lambda^{v}_{i}(\lambda^{v}_{i+1} - \lambda^{v}_{i})}. \end{aligned}$$

(b) *For TP*-6, *under Assumptions* A1 *and* A3:

$$
\begin{aligned}
\mathrm{UD}\max \mathrm{LR}_{2,T} &= \max_{1\le n_a\le N} n_a^{-1} \sup \mathrm{LR}^*_{2,T}(m_a, n_a, \varepsilon | n=0, m_a) \\
&\Rightarrow \max_{1\le n_a\le N} n_a^{-1} \sup_{(\lambda_1^v,\dots,\lambda_{n_a}^v)\in\Lambda_{v,\varepsilon}^c} \sum_{i=1}^{n_a} \frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v)\right)^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \\
&\le \max_{1\le n_a\le N} n_a^{-1} \sup_{(\lambda_1^v,\dots,\lambda_{n_a}^v)\in\Lambda_{v,\varepsilon}} \sum_{i=1}^{n_a} \frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v)\right)^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)}.
\end{aligned}
$$

(c) *For TP*-7, *under Assumptions* A2 *and* A3:

$$
\begin{aligned}
\mathrm{UD}\max \mathrm{LR}_{3,T} &= \max_{1\le m_a\le M} m_a^{-1} \sup \mathrm{LR}_{3,T}(m_a, n_a, \varepsilon | m=0, n_a) \\
&\Rightarrow \max_{1\le m_a\le M} m_a^{-1} \sup_{(\lambda_1^c,\dots,\lambda_{m_a}^c)\in\Lambda_{c,\varepsilon}^v} \sum_{j=1}^{m_a} \frac{\left\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\right\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\
&\le \max_{1\le m_a\le M} m_a^{-1} \sup_{(\lambda_1^c,\dots,\lambda_{m_a}^c)\in\Lambda_{c,\varepsilon}} \sum_{j=1}^{m_a} \frac{\left\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\right\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)}.
\end{aligned}
$$

(d) *For TP*-8, *under Assumptions* A1 *and* A2:

$$
\begin{aligned}
\mathrm{UD}\max \mathrm{LR}_{4,T} &= \max_{1\le n_a\le N} \max_{1\le m_a\le M} (n_a+m_a)^{-1} \sup \mathrm{LR}^*_{4,T}(m_a, n_a, \varepsilon | n=m=0) \\
&\Rightarrow \max_{1\le n_a\le N} \max_{1\le m_a\le M} (n_a+m_a)^{-1} \\
&\quad \times \sup_{(\lambda_1^c,\dots,\lambda_{m_a}^c;\lambda_1^v,\dots,\lambda_{n_a}^v)\in\Lambda_\varepsilon} \left[\sum_{j=1}^{m_a} \frac{\left\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\right\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \right. \\
&\quad \left. + \sum_{i=1}^{n_a} \frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v)\right)^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)}\right] \\
&\le \max_{1\le n_a\le N} \max_{1\le m_a\le M} (n_a+m_a)^{-1} \\
&\quad \times \sup_{(\lambda_1^c,\dots,\lambda_{m_a}^c;\lambda_1^v,\dots,\lambda_{n_a}^v)\in\Lambda_{cv,\varepsilon}} \left[\sum_{j=1}^{m_a} \frac{\left\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\right\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \right. \\
&\quad \left. + \sum_{i=1}^{n_a} \frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v)\right)^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)}\right].
\end{aligned}
$$

For TP-5 to TP-7, the critical values of the limit distributions are available in Bai and Perron (1998, 2003b) for N or M equal to 5. For TP-5 and TP-6, the results are valid for martingale differences or serially correlated errors. This is not the case for TP-7 and TP-8 for reasons discussed above. We then consider the maximum of the Wald-type tests discussed Section 4.2. The limit distribution applicable to TP-8 is new. Table 1 presents critical values obtained using simulations as discussed above for the case of a fixed number of breaks under H_1, for $\varepsilon = 0.1, 0.15$, and 0.20, and values of M and N up to 2; see Perron and Yamamoto (2019b) for additional critical values with $M, N = 2, 3, 4$.

4.4 *Testing for an additional break*

We now consider TP-9 and TP-10, which assess whether including an additional break is warranted. Let $(\tilde{T}_1^c, \ldots, \tilde{T}_m^c; \tilde{T}_1^v, \ldots, \tilde{T}_n^v)$ be the estimates of the break dates in δ and σ^2 obtained jointly by maximizing the quasi-likelihood function assuming m breaks in δ and n breaks in σ^2. For TP-9, the issue is whether an additional break in δ is present. The test is

$$\begin{aligned}\sup \mathrm{Seq}_{9,T}(m+1, n|m, n) = {} & \max_{1\le j\le m+1} \sup_{\tau \in \Lambda_{j,\varepsilon}^c} \log \hat{L}_T\big(\tilde{T}_1^c, \ldots, \tilde{T}_{j-1}^c, \tau, \tilde{T}_j^c, \ldots, \tilde{T}_m^c; \tilde{T}_1^v, \ldots, \tilde{T}_n^v\big) \\ & - \log \hat{L}_T\big(\tilde{T}_1^c, \ldots, \tilde{T}_m^c; \tilde{T}_1^v, \ldots, \tilde{T}_n^v\big),\end{aligned}$$

where $\Lambda_{j,\varepsilon}^c = \{\tau; \tilde{T}_{j-1}^c + (\tilde{T}_j^c - \tilde{T}_{j-1}^c)\varepsilon \le \tau \le \tilde{T}_j^c - (\tilde{T}_j^c - \tilde{T}_{j-1}^c)\varepsilon\}$. This amounts to performing $m+1$ tests for a single break in δ for each of the $m+1$ regimes defined by the partition $\{\tilde{T}_1^c, \ldots, \tilde{T}_m^c\}$. Note that there are different scenarios when allowing breaks in δ and in σ^2 to happen at different dates, since $(\tilde{T}_1^c, \ldots, \tilde{T}_m^c)$ and $(\tilde{T}_1^v, \ldots, \tilde{T}_n^v)$ can partly or completely overlap or be altogether different. This implies two possible cases: (1) if the n break dates in σ^2 are a subset of the m break dates in δ, there is no variance break between $\tilde{T}_{j-1}^c$ and $\tilde{T}_j^c$; (2) otherwise, there is one or more variance breaks between $\tilde{T}_{j-1}^c$ and $\tilde{T}_j^c$. In either cases, one can appeal to the results of Theorem 1(c) with $m_a = 1$ since any value of n_a is allowed, including 0. It is then easy to deduce that, in the case of martingale errors, the limit distribution of the test is, under Assumptions A2 and A3, $\lim_{T\to\infty} P(\sup \mathrm{Seq}_{9,T}(m+1, n|m, n) \le x) = G_{q,\varepsilon}(x)^{m+1}$, where $G_{q,\varepsilon}(x)$ is the cumulative distribution function of the random variable $\sup_{\lambda\in\Lambda_{1,\varepsilon}} \|(W_q(\lambda) - \lambda W_q(1))^2\| / (\lambda(1-\lambda))$, where $\Lambda_{1,\varepsilon} = \{\lambda; \varepsilon < \lambda < 1-\varepsilon\}$. The critical values of the distribution function $G_{q,\varepsilon}(x)^{m+1}$ can be found in Bai and Perron (1998, 2003b). With serial correlation in the errors, the principle is the same except that the statistic is based on the robust Wald test $\sup F_{3,T}$ as defined by (13) applied for a one break test to each segment. For TP-10, similar considerations apply. Here, the issue is whether an additional break in the variance is present. The test statistic is

$$\begin{aligned}& \sup \mathrm{Seq}_{10,T}(m, n+1|m, n) \\ & \quad = (2/\hat{\psi}) \max_{1\le i\le n+1} \sup_{\tau\in\Lambda_{i,\varepsilon}^v} \log \hat{L}_T\big(\tilde{T}_1^c, \ldots, \tilde{T}_m^c; \tilde{T}_1^v, \ldots, \tilde{T}_{i-1}^v, \tau, \tilde{T}_i^v, \ldots, \tilde{T}_m^v\big) \\ & \qquad - \log \hat{L}_T\big(\tilde{T}_1^c, \ldots, \tilde{T}_m^c; \tilde{T}_1^v, \ldots, \tilde{T}_n^v\big),\end{aligned}$$

where $\Lambda_{i,\varepsilon}^v = \{\tau; \tilde{T}_{i-1}^v + (\tilde{T}_i^v - \tilde{T}_{i-1}^v)\varepsilon \le \tau \le \tilde{T}_i^v - (\tilde{T}_i^v - \tilde{T}_{i-1}^v)\varepsilon\}$. The correction factor $(2/\hat{\psi})$ is needed to ensure that the limit distribution of the test is free of nuisance parameters when the errors are allowed to be nonnormal, serially correlated and conditionally heteroskedastic. One can then use part (b) of Theorem 1 to deduce that, under Assumptions A1 and A3 applied to each segments under H_0: $\lim_{T\to\infty} P(\sup \mathrm{Seq}_{10,T}(m, n+1|m, n) \leq x) = G_{1,\varepsilon}(x)^{n+1}$.

4.5 *Local asymptotic power*

Supplement D contains details about the local asymptotic power function of selected tests. We briefly summarize the relevant results. We consider model (1) focusing on the case of $n = m = 1$ with the following assumptions.

- Assumption L1. *Assumptions* A1 *and* A3 *hold with* $\sigma_{20} - \sigma_{10} = \sigma^*/\sqrt{T}$. *We also have* $T^{-1/2}\sum_{t=1}^{[Ts]}[(u_t^\sigma)^2 - 1] \Rightarrow \psi W(s)$ *with* $\psi = \lim_{T\to\infty} \mathrm{var}(T^{-1/2}\sum_{t=1}^{T}[(u_t^\sigma)^2 - 1])$ *and* $T^{-1}\sum_{t=1}^{[Ts]}(u_t^\sigma)^2 \xrightarrow{p} s$ *uniformly in* s.

- Assumption L2. *Assumptions* A2 *and* A3 *hold with* $\delta_2^0 - \delta_1^0 = \delta^*/\sqrt{T}$.

We derive the local asymptotic power of the tests $\sup \mathrm{LR}_{2,T}(n = 1, m = 1, \varepsilon|n = 0, m = 1)$ and $\sup \mathrm{LR}_{3,T}(m = 1, n = 1, \varepsilon|m = 0, n = 1)$ and the corresponding tests with no nuisance breaks accounted for, that is, $\sup \mathrm{LR}_{1,T}$ and the standard $\sup \mathrm{LR}_T$ test. Lemma S.1 shows that the local asymptotic power of the $\sup \mathrm{LR}_{2,T}$ test coincides with that of $\sup \mathrm{LR}_{1,T}$ except that the set of permissible break dates $\Lambda_{v,\epsilon}^c$ is smaller than $\Lambda_{v,\epsilon}$, which has no practical effect. Lemma S.2 shows that the local asymptotic power of $\sup \mathrm{LR}_{3,T}$ is the same as that of $\sup \mathrm{LR}_T$ derived in Andrews (1993, Theorem 4), again except that the set of permissible break dates is $\Lambda_{c,\epsilon}^v$ instead of $\Lambda_{c,\epsilon}$. Hence, when testing for changes in variance (resp., coefficients) allowing for changes in coefficients (resp., variance), we have the same local asymptotic power function as when testing for changes in variance (resp., coefficients) when no change in coefficient (resp., variance) is present. Hence, there is no loss in local asymptotic power adopting our more general approach.

We also derived the local asymptotic power function of the CUSQ test (see (14) below for its definition) and compared it to that of the $\sup \mathrm{LR}_{1,T}$ and $\sup \mathrm{LR}_{2,T}$ tests. Figure S.1 shows the asymptotic local power functions of the $\sup \mathrm{LR}_{1,T}$ and CUSQ tests when a break in variance occurs at $\lambda^{v0} = 0.3, 0.5$, and 0.7 and no break occurs in the coefficients. They show the local asymptotic power functions to be almost identical. Figure S.2 presents the local asymptotic power functions of the $\sup \mathrm{LR}_{2,T}$ test when it accounts for a coefficient break at $\lambda^{c0} = 0.3, 0.5$, or 0.7. It also shows that the local asymptotic power functions of the CUSQ test under the assumption of no break in the coefficients. This simulation design gives an advantage to the CUSQ. Indeed, the power of the $\sup \mathrm{LR}_{2,T}$ test is slightly lower when the variance and the coefficient break dates coincide. This is because the permissible break dates around the true break date are not considered due to the concurrent nuisance break. However, the power loss of the $\sup \mathrm{LR}_{2,T}$ test is very minor. The power of both tests are almost identical even though the $\sup \mathrm{LR}_{2,T}$ test considers a single nuisance break when the two breaks are far apart. that is, the case of $(\lambda^{v0}, \lambda^{c0}) = (0.3, 0.7)$ and $(0.7, 0.3)$.

5. Monte Carlo experiments

We provide simulation results to assess the size and power properties of some tests proposed; Section 5.1 for variance breaks, Section 5.2 for conditional tests, Section 5.3 for the $\sup \mathrm{LR}^*_{4,T}$ and $\mathrm{UD}\max$ tests. Supplement E provides additional results for the $\sup \mathrm{LR}_{1,T}$ and $\sup \mathrm{LR}_{2,T}$ tests with nonnormal errors. Following Bai and Ng (2005), we use: (a) the t distribution with 5 degrees of freedom, (b) a mixture of two normal distributions: $v_1 I(z \leq 0.5) + v_2 I(z > 0.5)$, where $z \sim U[0, 1]$, $v_1 \sim N(-1, 1)$ and $v_2 \sim N(1, 1)$, (c) the χ^2 distribution with 5 degrees of freedom and (d) an exponential distribution $-\ln(v)$, $v \sim U[0, 1]$. The results show that the exact size of the tests is similarly close to the nominal size. As expected, power is lower for all distributions, though the extent of the power loss is minor and the tests remain informative. Our tests for changes in variance retain their power advantage over the CUSQ test.

5.1 *Testing for variance breaks only*

We now consider the case of testing only for variance breaks assuming no change in δ. We investigate the properties of the following tests: the $\sup \mathrm{LR}^*_{1,T}(n_a, \varepsilon | m = n = 0)$, abbreviated $\sup \mathrm{LR}^*_{1,T}(n_a, \varepsilon)$ and the $\mathrm{UD}\max \mathrm{LR}_{1,T}$ for an unknown number of breaks up to $N = 5$. We also consider a corrected version of the CUSUM of squares test of Brown, Durbin, and Evans (1975), as extended by Deng and Perron (2008), given by

$$\mathrm{CUSQ} = \sup_{\lambda \in [0,1]} \left| T^{-1/2} \left[\sum_{t=1}^{[T\lambda]} \tilde{v}_t^2 - \big([T\lambda]/T\big) \sum_{t=1}^{T} \tilde{v}_t^2 \right] \right| / \hat{\varphi}_a^{1/2} \tag{14}$$

with $\hat{\varphi}_a = T^{-1} \sum_{j=-(T-1)}^{(T-1)} \omega(j, b_T) \sum_{t=|j|+1}^{T} \hat{\eta}_t \hat{\eta}_{t-j}$, where $\hat{\eta}_t = \tilde{v}_t^2 - \hat{\sigma}$, $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \tilde{v}_t^2$ and $\tilde{v}_t$ denotes the recursive residuals. Also $\omega(j, b_T)$ is the quadratic spectral kernel and the bandwidth b_T is selected using Andrews' (1991) method with an AR(1) approximation. The aim of the design is to address the following issues: (a) the size of the $\sup \mathrm{LR}^*_{1,T}(n_a, \varepsilon)$ and $\mathrm{UD}\max \mathrm{LR}_{1,T}$ tests; (b) the relative power of the three tests; (c) the power losses obtained when underspecifying the number of breaks; (d) the relative power of $\mathrm{UD}\max \mathrm{LR}_{1,T}$ compared to $\sup \mathrm{LR}^*_{1,T}(n_a, \varepsilon)$ with n_a specified to be the true number of breaks. We consider a dynamic model with GARCH errors, for which the DGP is given by $y_t = c + \alpha y_{t-1} + e_t$, $e_t = u_t \sqrt{h_t}$, $u_t \sim$ i.i.d. $N(0, 1)$, $h_t = \tau_1 + \tau_2 1(t > [0.5T]) + \gamma e_{t-1}^2 + \rho h_{t-1}$, where we set $h_0 = \tau_1/(1 - \gamma - \rho)$, $c = 0.5$, $\tau_1 = 0.1$, and $\varepsilon = 0.15$. We consider $\alpha = 0.2, 0.7$ and the GARCH$(1, 1)$ coefficients are set to $\gamma = 0.1, 0.3, 0.5$, and $\rho = 0.2$. The size and power of 5% nominal size tests are evaluated at $T = 100, 200$. The magnitude of the change τ_2 varies between 0 (size) and 0.3. The results are presented in Table 2. The $\sup \mathrm{LR}^*_{1,T}(1, \varepsilon)$ and $\mathrm{UD}\max \mathrm{LR}_{1,T}$ tests show size distortions when $\gamma = 0.5$ with $T = 100$ but the size is close to 5% when $T = 200$. The CUSQ test is slightly undersized. The $\mathrm{UD}\max \mathrm{LR}_{1,T}$ test has power close to that of $\sup \mathrm{LR}^*_{1,T}(1, \varepsilon)$, despite having a broader range of alternatives. The power of the latter two tests dominates that of CUSQ especially when $T = 100$. Supplement F shows the results to be robust for a static mean model with normal errors.

TABLE 2. Size and power of the $\sup \mathrm{LR}^*_{1,T}(n_a = 1, \varepsilon)$, $\mathrm{UD}\max \mathrm{LR}_{1,T}$ and CUSQ tests in a dynamic model with GARCH(1, 1) errors.

	$T = 100$																	
	$\alpha = 0.2$									$\alpha = 0.7$								
	$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$			$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$		
τ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.059	0.059	0.029	0.083	0.086	0.039	0.098	0.099	0.042	0.066	0.061	0.029	0.078	0.084	0.038	0.097	0.092	0.039
0.05	0.171	0.167	0.158	0.165	0.171	0.103	0.151	0.155	0.082	0.164	0.158	0.149	0.147	0.149	0.100	0.137	0.140	0.080
0.1	0.396	0.373	0.354	0.307	0.307	0.232	0.224	0.228	0.136	0.383	0.367	0.356	0.300	0.297	0.232	0.218	0.224	0.138
0.15	0.593	0.575	0.574	0.432	0.409	0.349	0.312	0.312	0.199	0.591	0.573	0.564	0.425	0.414	0.330	0.307	0.308	0.201
0.2	0.744	0.725	0.693	0.542	0.542	0.446	0.415	0.408	0.270	0.741	0.723	0.684	0.534	0.534	0.441	0.384	0.385	0.259
0.3	0.902	0.888	0.851	0.741	0.738	0.626	0.535	0.540	0.370	0.897	0.887	0.856	0.724	0.724	0.624	0.534	0.534	0.376
	$T = 200$																	
	$\alpha = 0.2$									$\alpha = 0.7$								
	$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$			$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$		
τ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.049	0.044	0.034	0.058	0.060	0.035	0.064	0.063	0.045	0.055	0.056	0.036	0.061	0.064	0.034	0.060	0.061	0.040
0.05	0.315	0.311	0.335	0.217	0.202	0.203	0.129	0.123	0.110	0.311	0.303	0.332	0.208	0.202	0.205	0.122	0.115	0.100
0.1	0.709	0.692	0.751	0.446	0.431	0.455	0.263	0.249	0.225	0.702	0.682	0.734	0.442	0.428	0.448	0.257	0.241	0.222
0.15	0.918	0.910	0.928	0.672	0.648	0.649	0.404	0.384	0.345	0.918	0.912	0.923	0.648	0.641	0.643	0.386	0.370	0.335
0.2	0.980	0.977	0.979	0.780	0.764	0.764	0.510	0.497	0.456	0.981	0.980	0.981	0.777	0.766	0.763	0.496	0.489	0.441
0.3	0.997	0.996	0.997	0.910	0.903	0.878	0.682	0.662	0.601	0.997	0.997	0.998	0.903	0.898	0.877	0.676	0.654	0.606

Note: DGP: $y_t = c + \alpha y_{t-1} + e_t$, $e_t = u_t\sqrt{h_t}$, with $u_t \sim$ i.i.d. $N(0, 1)$, $h_t = \tau_1 + \tau_2 1(t > [0.5T]) + \gamma e^2_{t-1} + \rho h_{t-1}$, $h_0 = \tau_1/(1 - \gamma - \rho)$, $c = 0.5$, $\tau_1 = 0.1$, $\rho = 0.2$; $\varepsilon = 0.15$.

We now turn to a case with two breaks in variance. The DGP is $y_t = e_t$; $e_t \sim$ i.i.d. $N(0, 1 + \theta 1(T^v_1 < t \leq T^v_2))$, that is, the variance increases at T^v_1 and returns to its original level at T^v_2. We consider two scenarios: $\{T^v_1 = [0.3T], T^v_2 = [0.6T]\}$ and $\{T^v_1 = [0.2T], T^v_2 = [0.8T]\}$. We set $T = 200$ and $\varepsilon = 0.10, 0.15$. The magnitude of the break in σ^2 varies between $\theta = 0$ (size) and $\theta = 3$. We again consider the $\mathrm{UD}\max \mathrm{LR}_{1,T}$ test with $N = 5$ but include both the $\sup \mathrm{LR}^*_{1,T}(1, \varepsilon)$ test for a single break and the $\sup \mathrm{LR}^*_{1,T}(2, \varepsilon)$ test for two breaks to assess the extent of power gains when specifying the correct number of breaks. The results are presented in Table 3. Consider first the size of the tests. The $\sup \mathrm{LR}^*_{1,T}(1, \varepsilon)$, $\sup \mathrm{LR}^*_{1,T}(2, \varepsilon)$ and $\mathrm{UD}\max \mathrm{LR}_{1,T}$ tests are slightly conservative and the CUSQ even more so with an exact size of 0.025. As expected, power increases as ε increases since the range of alternatives is smaller. When comparing the $\sup \mathrm{LR}^*_{1,T}(1, \varepsilon)$ and $\sup \mathrm{LR}^*_{1,T}(2, \varepsilon)$ tests, the latter is more powerful, indicating that allowing for the correct number of breaks improves power. The $\mathrm{UD}\max \mathrm{LR}_{1,T}$ test has power between those of the $\sup \mathrm{LR}^*_{1,T}(1, \varepsilon)$ and $\sup \mathrm{LR}^*_{1,T}(2, \varepsilon)$ tests. These tests are considerably more powerful than the CUSQ, which has little power.

5.2 *Conditional tests*

We now consider the properties of the tests that condition on either breaks in coefficients (resp., variance) when testing for changes in variance (resp., coefficients). Consider first the size and power of $\sup \mathrm{LR}^*_{2,T}(m_a, n_a, \varepsilon | n = 0, m_a)$ which tests for n_a changes in σ^2 conditional on m_a changes in δ with $\varepsilon = 0.1, 0.2$. We set $m_a = n_a = 1$ and the DGP is

TABLE 3. Size and power of the $\sup \mathrm{LR}^*_{1,T}(n_a, \varepsilon)$, $\mathrm{UD}\max \mathrm{LR}_{1,T}$ and CUSQ tests with normal errors and two variance breaks.

	$T_1^v=[0.3T]$, $T_2^v=[0.6T]$							$T_1^v=[0.2T]$, $T_2^v=[0.8T]$						
	$\varepsilon=0.10$			$\varepsilon=0.15$				$\varepsilon=0.10$			$\varepsilon=0.15$			
θ	$n_a=1$	$n_a=2$	UDmax	$n_a=1$	$n_a=2$	UDmax	CUSQ	$n_a=1$	$n_a=2$	UDmax	$n_a=1$	$n_a=2$	UDmax	CUSQ
0	0.035	0.034	0.036	0.033	0.025	0.030	0.025	0.035	0.034	0.036	0.033	0.025	0.030	0.025
0.25	0.049	0.040	0.045	0.066	0.054	0.064	0.031	0.067	0.043	0.062	0.063	0.052	0.064	0.035
0.5	0.111	0.120	0.103	0.117	0.159	0.121	0.059	0.158	0.138	0.139	0.166	0.170	0.165	0.036
0.75	0.164	0.260	0.195	0.171	0.294	0.209	0.085	0.263	0.283	0.265	0.276	0.360	0.287	0.044
1	0.213	0.418	0.289	0.239	0.493	0.340	0.124	0.390	0.472	0.390	0.428	0.520	0.442	0.061
1.25	0.291	0.575	0.404	0.328	0.674	0.495	0.147	0.538	0.647	0.558	0.563	0.707	0.606	0.053
1.5	0.356	0.703	0.513	0.405	0.778	0.613	0.197	0.647	0.780	0.676	0.706	0.837	0.731	0.065
2	0.456	0.835	0.701	0.530	0.893	0.761	0.276	0.798	0.915	0.841	0.828	0.946	0.868	0.083
2.5	0.621	0.935	0.848	0.686	0.959	0.882	0.375	0.907	0.971	0.931	0.930	0.986	0.950	0.133
3	0.693	0.959	0.895	0.728	0.983	0.919	0.430	0.943	0.987	0.961	0.963	0.993	0.977	0.120

Note: DGP: $y_t = e_t$; $e_t \sim$ i.i.d. $N(0, 1+\theta 1(T_1^v < t \le T_2^v))$, $T=200$.

a simple mean shift model with a change of magnitude μ_2 at mid-sample with i.i.d. normal errors having a change in variance of magnitude θ (under H_1) that occurs at $[0.25T]$. The results for size are presented in Table 4. The test is slightly conservative and more so as the trimming is larger. This is due to the fact that the limit distribution used is an upper bound. The results for power are presented in Table 5. It increases rapidly with the magnitude of the variance break θ and with T. It also marginally increases with the value of the trimming ε.

We next investigate the size and power of $\sup \mathrm{LR}^*_{3,T}(m_a, n_a, \varepsilon | m=0, n_a)$ which tests for m_a changes in δ conditional on n_a changes in σ^2 with $\varepsilon = 0.1, 0.2$. We again set $m_a = n_a = 1$ and consider the mean model in which σ^2 changes at mid-sample. We also consider an AR(1) model $y_t = c + \alpha y_{t-1} + e_t$ with $c = 0.5$, $\alpha = 0.5$ and e_t being i.i.d. normal

TABLE 4. Size of the $\sup \mathrm{LR}^*_{2,T}(m_a=1, n_a=1, \varepsilon | n=0, m_a=1)$ test with different trimming parameter ε in the case of normal errors.

	$T=100$				$T=200$			
$\mu_2 \setminus \varepsilon$	0.1	0.15	0.2	0.25	0.1	0.15	0.2	0.25
0	0.045	0.042	0.030	0.023	0.039	0.032	0.030	0.031
0.1	0.038	0.028	0.033	0.030	0.045	0.046	0.036	0.037
0.25	0.037	0.039	0.034	0.030	0.034	0.034	0.035	0.030
0.5	0.037	0.035	0.036	0.033	0.031	0.025	0.029	0.027
0.75	0.043	0.047	0.046	0.041	0.044	0.033	0.035	0.031
1	0.034	0.031	0.031	0.031	0.034	0.027	0.020	0.017
2	0.030	0.023	0.028	0.028	0.041	0.029	0.028	0.029
4	0.034	0.032	0.031	0.027	0.034	0.026	0.024	0.026
10	0.038	0.033	0.032	0.031	0.038	0.033	0.025	0.022
20	0.031	0.030	0.035	0.027	0.040	0.034	0.023	0.021

Note: DGP: $y_t = \mu_1 + \mu_2 1(t > [0.5T]) + e_t$, $e_t \sim$ i.i.d. $N(0,1)$, $\mu_1 = 0$.

TABLE 5. Power of the $\sup \mathrm{LR}^*_{2,T}(m_a = 1, n_a = 1, \varepsilon | n = 0, m_a = 1)$ test with different trimming parameter ε in the case of normal errors.

	$T = 100$													
	$\varepsilon = 0.1$							$\varepsilon = 0.2$						
$\theta \setminus \mu_2$	0	0.1	0.5	2	4	10	20	0	0.1	0.5	2	4	10	20
0.25	0.063	0.046	0.047	0.056	0.065	0.063	0.053	0.056	0.040	0.043	0.049	0.045	0.044	0.047
0.5	0.101	0.094	0.089	0.090	0.099	0.096	0.101	0.091	0.092	0.097	0.077	0.096	0.091	0.101
0.75	0.150	0.162	0.133	0.168	0.177	0.181	0.178	0.168	0.174	0.160	0.176	0.177	0.176	0.171
1	0.237	0.233	0.218	0.212	0.222	0.244	0.242	0.270	0.285	0.226	0.225	0.231	0.236	0.235
1.25	0.270	0.300	0.319	0.293	0.353	0.362	0.327	0.318	0.323	0.335	0.316	0.375	0.383	0.321
1.5	0.388	0.379	0.378	0.419	0.417	0.448	0.398	0.443	0.431	0.435	0.425	0.448	0.462	0.445
2	0.533	0.519	0.496	0.557	0.556	0.598	0.559	0.592	0.586	0.558	0.588	0.602	0.620	0.594
3	0.760	0.771	0.771	0.779	0.830	0.843	0.802	0.827	0.823	0.825	0.822	0.857	0.863	0.838
4	0.887	0.876	0.865	0.892	0.908	0.909	0.916	0.921	0.910	0.920	0.924	0.927	0.943	0.940
	$T = 200$													
	$\varepsilon = 0.1$							$\varepsilon = 0.2$						
$\theta \setminus \mu_2$	0	0.1	0.5	2	4	10	20	0	0.1	0.5	2	4	10	20
0.25	0.052	0.066	0.066	0.077	0.084	0.092	0.090	0.063	0.067	0.059	0.073	0.074	0.067	0.071
0.5	0.175	0.177	0.153	0.204	0.178	0.207	0.219	0.205	0.188	0.165	0.216	0.185	0.199	0.212
0.75	0.311	0.352	0.340	0.361	0.382	0.369	0.365	0.383	0.385	0.364	0.376	0.384	0.385	0.381
1	0.485	0.506	0.469	0.518	0.553	0.529	0.567	0.551	0.566	0.529	0.542	0.585	0.574	0.599
1.25	0.648	0.643	0.660	0.716	0.716	0.717	0.741	0.695	0.685	0.694	0.729	0.745	0.760	0.770
1.5	0.771	0.771	0.773	0.821	0.827	0.842	0.821	0.834	0.813	0.824	0.852	0.851	0.871	0.851
2	0.918	0.907	0.928	0.933	0.962	0.942	0.955	0.943	0.943	0.953	0.950	0.972	0.961	0.973
3	0.990	0.996	0.992	0.996	0.999	0.998	0.996	0.997	0.998	0.996	0.996	0.999	0.999	0.998
4	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: DGP: $y_t = \mu_1 + \mu_2 1(t > [0.5T]) + e_t$, $e_t \sim$ i.i.d. $N(0, 1 + \theta 1(t > [0.25T]))$.

errors having a change in variance at $[0.5T]$ with magnitude θ. This is done to investigate potential size distortions due to large variance changes. As discussed in Section 4.1, a change in variance induces a change in the marginal distribution of the regressors when lagged dependent variables are included. The results for the size of the tests are presented in Table 6. The size under the mean model is close to the nominal level but the test becomes conservative as ε increases since the limiting distribution used is a bound. The size under the AR(1) model is very similar with the distortions being even smaller. This indicates that the shrinking variance assumption is not binding. The results for power are presented in Table 7 for the mean model with a coefficient change at $[0.25T]$. The power quickly increases as the break magnitude θ and T increase. The power again marginally increases with ε.

5.3 *Size and power of the* $\sup \mathrm{LR}^*_{4,T}$ *and* $\mathrm{UD}\max \mathrm{LR}_{4,T}$ *tests*

We now consider the $\sup \mathrm{LR}^*_{4,T}$ and $\mathrm{UD}\max \mathrm{LR}_{4,T}$ (simply labeled UD max) tests. To this end, we use a model with GARCH(1, 1) errors so that the DGP is $y_t = e_t$ with $e_t = u_t\sqrt{h_t}$, where $u_t \sim$ i.i.d. $N(0, 1)$, $h_t = \tau_1 + \gamma e_{t-1}^2 + \rho h_{t-1}$, $h_0 = \tau_1/(1 - \gamma - \rho)$, $\tau_1 = 1$, $\rho = 0.2$ and γ

TABLE 6. Size of the $\sup \mathrm{LR}^*_{3,T}(m_a = 1, n_a = 1, \varepsilon | m = 0, n_a = 1)$ test with different trimming parameter ε in the case of normal errors.

	$T = 100$				$T = 200$			
$\theta \setminus \varepsilon$	0.1	0.15	0.2	0.25	0.1	0.15	0.2	0.25
				Panel (a)				
0	0.043	0.053	0.051	0.031	0.042	0.041	0.039	0.036
0.1	0.050	0.053	0.033	0.037	0.027	0.035	0.033	0.026
0.25	0.042	0.042	0.042	0.023	0.034	0.044	0.039	0.040
0.5	0.044	0.024	0.038	0.038	0.036	0.035	0.035	0.028
0.75	0.039	0.039	0.037	0.033	0.043	0.038	0.040	0.034
1	0.033	0.043	0.045	0.027	0.029	0.044	0.042	0.029
2	0.046	0.045	0.039	0.022	0.038	0.032	0.029	0.013
4	0.030	0.054	0.035	0.020	0.038	0.032	0.030	0.014
10	0.034	0.043	0.030	0.027	0.037	0.035	0.031	0.015
20	0.046	0.039	0.027	0.027	0.032	0.039	0.030	0.012
				Panel (b)				
0	0.069	0.066	0.066	0.055	0.049	0.043	0.050	0.042
0.1	0.057	0.060	0.062	0.056	0.044	0.047	0.048	0.039
0.25	0.057	0.055	0.055	0.049	0.039	0.044	0.053	0.035
0.5	0.050	0.058	0.048	0.043	0.051	0.044	0.050	0.035
0.75	0.055	0.055	0.057	0.046	0.043	0.036	0.036	0.034
1	0.065	0.055	0.051	0.042	0.044	0.053	0.045	0.028
2	0.047	0.066	0.062	0.045	0.043	0.040	0.040	0.027
4	0.052	0.053	0.039	0.025	0.030	0.051	0.031	0.017
10	0.050	0.063	0.050	0.026	0.043	0.038	0.034	0.018
20	0.040	0.065	0.059	0.024	0.048	0.038	0.034	0.025

Note: Panel (a): DGP: $y_t = \mu_1 + e_t$, $e_t \sim$ i.i.d. $N(0, 1 + \theta 1(t > [0.5T]))$, $\mu_1 = 0$; panel (b): DGP: $y_t = c + \alpha y_{t-1} + e_t$, $e_t \sim$ i.i.d. $N(0, 1 + \theta 1(t > [0.5T]))$, $c = 0$, $\alpha = 0.5$.

takes values $0.1, 0.3, 0.5$. Also, $\varepsilon = 0.1, 0.2$. For the UD max test, $M = N = 2$ and for the $\sup \mathrm{LR}^*_{4,T}$ test, we consider the following combinations: (a) $m_a = n_a = 1$, (b) $m_a = 1$, $n_a = 2$, (c) $m_a = 2$, $n_a = 1$. We set $T = 100, 200$. The results, presented in Table 8, show that the size is close to or slightly lower than the nominal 5% level (some cases have slight liberal size distortions when $T = 100$, which, however, decrease when $T = 200$). Supplement G shows that the tests have good sizes with i.i.d. normal errors.

We now consider the power of these tests. Since some partial results for the one break case are available in Tables S.6–S.7 for the $\sup \mathrm{LR}^*_{4,T}$ test, we concentrate on the case with a different number of breaks in coefficients and in variance. We also only consider i.i.d. normal errors though the hybrid-type correction is still applied. Table 9 presents the results for the case with one break in coefficient and two breaks in variance, in which case the DGP is $y_t = \mu_1 + \mu_2 1(t > T^c) + e_t$, $e_t \sim$ i.i.d. $N(0, 1 + \theta 1(T_1^v < t \leq T_2^v))$ with $\mu_1 = 0$, $\mu_2 = \theta$ and $\varepsilon = 0.1$. Five different configurations of break dates are considered. We analyze two forms of the $\sup \mathrm{LR}^*_{4,T}$ test: (a) one testing for a single break in both mean and variance, (b) one correctly testing for two changes in variance and one change in mean. This is done to investigate the extent of the power differences when underspecifying the number of breaks. As expected, the power increases rapidly with θ and with T. With the DGP used, the power is similar whether accounting for one or (correctly) two breaks in

TABLE 7. Power of the $\sup \mathrm{LR}^*_{3,T}(m_a = 1, n_a = 1, \varepsilon | m = 0, n_a = 1)$ test with different trimming parameter ε in the case of normal errors.

	$T = 100$													
	$\varepsilon = 0.1$							$\varepsilon = 0.2$						
$\mu_2 \setminus \theta$	0	0.1	0.5	2	4	10	20	0	0.1	0.5	2	4	10	20
0.1	0.050	0.050	0.055	0.058	0.059	0.057	0.059	0.050	0.049	0.043	0.034	0.031	0.037	0.030
0.25	0.096	0.092	0.092	0.082	0.078	0.074	0.080	0.117	0.115	0.110	0.088	0.077	0.077	0.077
0.5	0.349	0.351	0.340	0.300	0.263	0.255	0.245	0.353	0.350	0.334	0.305	0.283	0.283	0.243
0.75	0.670	0.663	0.651	0.580	0.538	0.503	0.485	0.702	0.696	0.692	0.625	0.586	0.586	0.544
1	0.901	0.899	0.892	0.853	0.821	0.799	0.785	0.930	0.929	0.929	0.901	0.866	0.866	0.811
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$T = 200$													
	$\varepsilon = 0.1$							$\varepsilon = 0.2$						
$\mu_2 \setminus \theta$	0	0.1	0.5	2	4	10	20	0	0.1	0.5	2	4	10	20
0.1	0.059	0.062	0.054	0.046	0.043	0.045	0.049	0.059	0.056	0.044	0.058	0.055	0.053	0.042
0.25	0.175	0.170	0.178	0.140	0.136	0.136	0.138	0.192	0.179	0.183	0.158	0.142	0.132	0.135
0.5	0.650	0.609	0.585	0.556	0.518	0.494	0.466	0.681	0.655	0.673	0.583	0.542	0.506	0.482
0.75	0.939	0.959	0.934	0.913	0.901	0.882	0.847	0.963	0.965	0.963	0.913	0.909	0.878	0.883
1	1.000	0.999	0.997	0.995	0.989	0.988	0.987	1.000	0.998	0.999	0.998	0.998	0.996	0.995
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: DGP: $y_t = \mu_1 + \mu_2 1(t > [0.25T]) + e_t$, $e_t \sim$ i.i.d. $N(0, 1 + \theta 1(t > [0.5T]))$, $\mu_1 = 0$.

variance and the power of the UD max test is also similar to the power of both versions of the $\sup \mathrm{LR}^*_{4,T}$ test. This may, however, be DGP specific. Table 10 presents the results for the case with two breaks in coefficient and one break in variance, with the DGP given by $y_t = \mu_1 + \mu_2 1(T^c_1 < t \leq T^c_2) + e_t$, $e_t \sim$ i.i.d. $N(0, 1 + \theta 1(t > T^v))$ with $\mu_1 = 0$ and $\mu_2 = \theta$.

TABLE 8. Size of the $\sup \mathrm{LR}^*_{4,T}(m_a, n_a)$ and $\mathrm{UD}\max \mathrm{LR}_{4,T}$ tests in the case of GARCH(1, 1) errors.

	$T = 100$							
	$\varepsilon = 0.1$				$\varepsilon = 0.2$			
γ	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax
0.1	0.044	0.046	0.047	0.050	0.037	0.040	0.035	0.046
0.3	0.048	0.065	0.051	0.073	0.041	0.052	0.042	0.055
0.5	0.072	0.083	0.075	0.085	0.065	0.069	0.059	0.061
	$T = 200$							
	$\varepsilon = 0.1$				$\varepsilon = 0.2$			
γ	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax
0.1	0.034	0.035	0.034	0.041	0.036	0.034	0.037	0.037
0.3	0.032	0.041	0.035	0.043	0.036	0.037	0.031	0.040
0.5	0.039	0.044	0.041	0.051	0.040	0.040	0.024	0.040

Note: DGP: $y_t = e_t$, $e_t = u_t\sqrt{h_t}$, with $u_t \sim$ i.i.d. $N(0, 1)$, $h_t = \tau_1 + \gamma e^2_{t-1} + \rho h_{t-1}$, $\tau_1 = 1$, $\rho = 0.2$, $h_0 = \tau_1/(1 - \gamma - \rho)$.

Table 9. Power of the $\sup \mathrm{LR}^*_{4,T}(m_a, n_a)$ and $\mathrm{UD}\max \mathrm{LR}_{4,T}$ tests for DGPs with one break in coefficients and two breaks in variance.

	$T = 100$														
	$T^c = T_1^v = [0.3T]$, $T_2^v = [0.6T]$			$T^c = T_2^v = [0.6T]$, $T_1^v = [0.3T]$			$T^c = [0.3T]$, $T_1^v = [0.5T]$, $T_2^v = [0.6T]$			$T^c = [0.5T]$, $T_1^v = [0.3T]$, $T_2^v = [0.6T]$			$T^c = [0.6T]$, $T_1^v = [0.3T]$, $T_2^v = [0.5T]$		
θ	$m_a = 1$ $n_a = 1$	$m_a = 1$ $n_a = 2$	UDmax	$m_a = 1$ $n_a = 1$	$m_a = 1$ $n_a = 2$	UDmax	$m_a = 1$ $n_a = 1$	$m_a = 1$ $n_a = 2$	UDmax	$m_a = 1$ $n_a = 1$	$m_a = 1$ $n_a = 2$	UDmax	$m_a = 1$ $n_a = 1$	$m_a = 1$ $n_a = 2$	UDmax
0.25	0.081	0.069	0.090	0.091	0.082	0.097	0.083	0.069	0.086	0.089	0.085	0.097	0.092	0.079	0.100
0.5	0.263	0.263	0.280	0.314	0.280	0.313	0.262	0.233	0.269	0.320	0.294	0.326	0.318	0.281	0.315
0.75	0.576	0.560	0.586	0.655	0.631	0.643	0.592	0.570	0.583	0.687	0.661	0.691	0.648	0.628	0.650
1	0.854	0.860	0.857	0.892	0.902	0.896	0.874	0.861	0.877	0.895	0.906	0.918	0.890	0.886	0.888
1.25	0.980	0.974	0.976	0.988	0.985	0.984	0.982	0.974	0.982	0.986	0.983	0.987	0.983	0.987	0.987
1.5	1.000	1.000	0.997	0.998	0.999	1.000	0.999	0.997	0.998	1.000	1.000	1.000	1.000	0.999	0.999
	$T = 200$														
	$T^c = T_1^v = [0.3T]$, $T_2^v = [0.6T]$			$T^c = T_2^v = [0.6T]$, $T_1^v = [0.3T]$			$T^c = [0.3T]$, $T_1^v = [0.5T]$, $T_2^v = [0.6T]$			$T^c = [0.5T]$, $T_1^v = [0.3T]$, $T_2^v = [0.6T]$			$T^c = [0.6T]$, $T_1^v = [0.3T]$, $T_2^v = [0.5T]$		
θ	$m_a = 1$ $n_a = 1$	$m_a = 1$ $n_a = 2$	UDmax	$m_a = 1$ $n_a = 1$	$m_a = 1$ $n_a = 2$	UDmax	$m_a = 1$ $n_a = 1$	$m_a = 1$ $n_a = 2$	UDmax	$m_a = 1$ $n_a = 1$	$m_a = 1$ $n_a = 2$	UDmax	$m_a = 1$ $n_a = 1$	$m_a = 1$ $n_a = 2$	UDmax
0.25	0.119	0.124	0.129	0.156	0.138	0.159	0.128	0.109	0.125	0.152	0.153	0.158	0.142	0.134	0.149
0.5	0.552	0.561	0.569	0.633	0.622	0.637	0.547	0.515	0.545	0.642	0.645	0.656	0.628	0.593	0.624
0.75	0.925	0.929	0.925	0.961	0.958	0.955	0.935	0.927	0.931	0.968	0.976	0.971	0.966	0.956	0.962
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.999	1.000
1.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: DGP: $y_t = \mu_1 + \mu_2 1(t > T^c) + e_t$, $e_t \sim$ i.i.d. $N(0, 1 + \theta 1(T_1^v < t \leq T_2^v))$, $\mu_1 = 0$, $\mu_2 = \theta$, $\varepsilon = 0.1$.

TABLE 10. Power of the $\sup \mathrm{LR}^*_{4,T}(m_a, n_a)$ and $\mathrm{UD}\max \mathrm{LR}_{4,T}$ tests for DGPs with two breaks in coefficients and one break in variance.

	$T=100$														
	$T_1^c = T^v = [0.3T]$, $T_2^c = [0.6T]$			$T_1^c = [0.3T]$, $T_2^c = T^v = [0.6T]$			$T_1^c = [0.5T]$, $T_2^c = [0.6T]$, $T^v = [0.3T]$			$T_1^c = [0.3T]$, $T_2^c = [0.6T]$, $T^v = [0.5T]$			$T_1^c = [0.3T]$, $T_2^c = [0.5T]$, $T^v = [0.6T]$		
	$m_a=1$	$m_a=2$		$m_a=1$	$m_a=2$		$m_a=1$	$m_a=2$		$m_a=1$	$m_a=2$		$m_a=1$	$m_a=2$	
θ	$n_a=1$	$n_a=1$	UDmax	$n_a=1$	$n_a=1$	UDmax	$n_a=1$	$n_a=1$	UDmax	$n_a=1$	$n_a=1$	UDmax	$n_a=1$	$n_a=1$	UDmax
0.25	0.064	0.085	0.081	0.076	0.080	0.091	0.051	0.051	0.056	0.073	0.080	0.087	0.061	0.065	0.080
0.5	0.107	0.181	0.164	0.140	0.194	0.177	0.085	0.092	0.101	0.141	0.194	0.178	0.098	0.150	0.134
0.75	0.238	0.405	0.352	0.294	0.479	0.423	0.137	0.161	0.165	0.282	0.468	0.416	0.187	0.358	0.307
1	0.412	0.688	0.612	0.512	0.809	0.750	0.237	0.278	0.276	0.498	0.770	0.720	0.302	0.677	0.577
1.25	0.566	0.849	0.800	0.709	0.958	0.923	0.299	0.436	0.405	0.684	0.933	0.905	0.399	0.876	0.806
1.5	0.707	0.935	0.902	0.854	0.995	0.987	0.409	0.564	0.558	0.835	0.987	0.980	0.520	0.964	0.934
	$T=200$														
	$T_1^c = T^v = [0.3T]$, $T_2^c = [0.6T]$			$T_1^c = [0.3T]$, $T_2^c = T^v = [0.6T]$			$T_1^c = [0.5T]$, $T_2^c = [0.6T]$, $T^v = [0.3T]$			$T_1^c = [0.3T]$, $T_2^c = [0.6T]$, $T^v = [0.5T]$			$T_1^c = [0.3T]$, $T_2^c = [0.5T]$, $T^v = [0.6T]$		
	$m_a=1$	$m_a=2$		$m_a=1$	$m_a=2$		$m_a=1$	$m_a=2$		$m_a=1$	$m_a=2$		$m_a=1$	$m_a=2$	
θ	$n_a=1$	$n_a=1$	UDmax	$n_a=1$	$n_a=1$	UDmax	$n_a=1$	$n_a=1$	UDmax	$n_a=1$	$n_a=1$	UDmax	$n_a=1$	$n_a=1$	UDmax
0.25	0.079	0.117	0.097	0.083	0.104	0.113	0.060	0.066	0.073	0.086	0.115	0.114	0.075	0.088	0.096
0.5	0.275	0.421	0.379	0.330	0.490	0.447	0.145	0.180	0.185	0.335	0.486	0.453	0.224	0.378	0.340
0.75	0.590	0.814	0.774	0.688	0.913	0.870	0.333	0.420	0.408	0.681	0.895	0.852	0.480	0.818	0.769
1	0.844	0.976	0.963	0.919	0.997	0.994	0.524	0.673	0.654	0.891	0.992	0.986	0.700	0.982	0.960
1.25	0.964	0.996	0.992	0.991	1.000	1.000	0.736	0.868	0.859	0.983	1.000	1.000	0.861	0.999	0.999
1.5	0.995	1.000	1.000	1.000	1.000	1.000	0.853	0.945	0.942	0.999	1.000	1.000	0.944	1.000	1.000

Note: DGP: $y_t = \mu_1 + \mu_2 1(T_1^c < t \le T_2^c) + e_t$, $e_t \sim$ i.i.d. $N(0, 1 + \theta 1(t > T^v))$, $\mu_1 = 0$, $\mu_2 = \theta$, $\varepsilon = 0.1$.

Again, we consider two forms of the $\sup \mathrm{LR}^*_{4,T}$ test: one testing for a single break in both mean and variance, one correctly testing for two changes in mean and one change in variance. Table 10 shows that for given values of θ and T, the power is lower than with one break in coefficient and two breaks in variance. Also, the UD max test now has power between that of the test correctly specifying the type and number of breaks and that underspecifying the number of changes in mean. The difference can be substantial and, as in Bai and Perron (2006), the power of the UD max test is close to that attainable when the type and number of breaks is correctly specified.

6. Estimating the numbers of breaks in coefficients and in variance

To select the number of breaks in regression coefficients or error variance, we suggest a specific to general procedure that uses the sequential tests proposed in Section 4.4. We determine the number of coefficients and variance breaks allowing for a given number of breaks in the other component. When selecting the number of breaks in δ, we consider TP-9 and the test $\sup \mathrm{Seq}_{9,T}(m+1, N|m, N)$ is applied, starting with $H_0: m = 0$ and $H_1: m = 1$, where N is some prespecified maximum number of breaks in variance. Upon a rejection, we proceed to $H_0: m = 1$ versus $H_1: m = 2$, and so on until the test stops rejecting. Since the number of breaks n in σ^2 is unknown, contamination of the test statistics by unaccounted breaks in σ^2 must be avoided. This can be achieved imposing a maximum number N throughout. Similarly, to select the number of breaks in σ^2, TP-10 is considered and the test $\sup \mathrm{Seq}_{10,T}(M, n+1|M, n)$ is used for $n = 0, 1, \ldots,$ until a nonrejection occurs. Again, some maximum number of breaks in the coefficients M is imposed. We performed a simple simulation experiment with $T = 200$, $\varepsilon = 0.15$ and the DGP given by

$$y_t = \mu_1 + \mu_2 1(t > T^c) + e_t, \quad e_t \sim \text{i.i.d. } N\big(0, 1 + \theta 1(t > T^v)\big)$$

with $\mu_1 = 0$ so that at most one break in either mean or variance occurs. We consider the following scenarios: (a) no change in mean or variance, (b) a change in mean only occurring at mid-sample, (c) a change in variance only occurring at mid-sample, (d) a change in both mean and variance occurring at a common date (mid-sample); (e) a change in both mean and variance occurring at different but close dates ($T^c = [0.5T]$, $T^v = [0.7T]$) or (f) at different and distant dates ($T^c = [0.25T]$, $T^v = [0.75T]$). Different magnitudes of breaks are considered. The procedure is applied setting the maximum number of breaks to $M = 2$ and $N = 2$ (i.e., four breaks overall). We also considered a split-sample method discussed in Supplement H. The results are presented in Tables 11 and S.4. The procedures work quite well in selecting the correct number and type of breaks. There are cases, however, where the probability of correct selection is quite low with the split-sample method, for example, when both changes in mean and variance are not large and occur at different dates, especially far apart. The specific to general approach tests for breaks in coefficients and variance separately allowing the other component to have unknown breaks, which can avoid segmentations and lead to power gains. The probabilities of selecting the correct number of each type of breaks are high with this approach (higher than with the split-sample method, see Table S.10) when the changes are not large and the break dates are different. Hence, we recommend this procedure in practice.

TABLE 11. Finite sample performance of the specific to general sequential procedure to select the number of breaks in coefficients and variance.

	$m=n=0$	$m=n=1$ $T^c=[0.5T]$, $T^v=[0.7T]$				$m=n=1$ $T^c=[0.25T]$, $T^v=[0.75T]$		
		$\mu_2=\theta=1$	$\mu_2=1,\theta=3$	$\mu_2=1,\theta=5$	$\mu_2=\theta=2$	$\mu_2=\theta=1$	$\mu_2=\theta=2$	$\mu_2=1,\theta=3$
prob$(m=0,n=0)$	**0.906**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
prob$(m=0,n=1)$	0.042	0.000	0.002	0.003	0.000	0.000	0.000	0.000
prob$(m=0,n=2)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
prob$(m=1,n=0)$	0.043	0.286	0.001	0.000	0.023	0.343	0.028	0.004
prob$(m=1,n=1)$	0.007	**0.680**	**0.954**	**0.956**	**0.936**	**0.628**	**0.937**	**0.963**
prob$(m=1,n=2)$	0.000	0.009	0.002	0.016	0.019	0.007	0.011	0.010
prob$(m=2,n=0)$	0.002	0.008	0.000	0.000	0.000	0.010	0.001	0.000
prob$(m=2,n=1)$	0.000	0.016	0.023	0.025	0.022	0.011	0.020	0.021
prob$(m=2,n=2)$	0.000	0.001	0.000	0.000	0.000	0.001	0.003	0.002

	$m=n=1$ $T^c=T^v=[0.5T]$		$m=1,n=0$ $T^c=[0.5T]$			$m=0,n=1$ $T^v=[0.5T]$		
	$\mu_2=\theta=1$	$\mu_2=1,\theta=3$	$\mu_2=1$	$\mu_2=2$	$\mu_2=3$	$\theta=1$	$\theta=2$	$\theta=3$
prob$(m=0,n=0)$	0.000	0.000	0.000	0.000	0.000	0.234	0.005	0.000
prob$(m=0,n=1)$	0.003	0.029	0.000	0.000	0.000	**0.706**	**0.924**	**0.924**
prob$(m=0,n=2)$	0.000	0.002	0.000	0.000	0.000	0.013	0.027	0.031
prob$(m=1,n=0)$	0.240	0.000	**0.931**	**0.935**	**0.934**	0.009	0.000	0.000
prob$(m=1,n=1)$	**0.729**	**0.917**	0.039	0.038	0.038	0.035	0.040	0.041
prob$(m=1,n=2)$	0.008	0.034	0.000	0.000	0.000	0.002	0.003	0.003
prob$(m=2,n=0)$	0.005	0.000	0.028	0.023	0.024	0.001	0.000	0.000
prob$(m=2,n=1)$	0.014	0.017	0.002	0.004	0.004	0.000	0.001	0.001
prob$(m=2,n=2)$	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000

Note: DGP: $y_t=\mu_1+\mu_2 1(t>T^c)+e_t$, $e_t\sim$ i.i.d. $N(0,1+\theta 1(t>T^v))$, $\varepsilon=0.15$, $T=200$. prob$(m=j,n=i)$ represents the probability of choosing j breaks in mean and i breaks in variance. The upper bounds for the coefficients and the variance breaks are set to $M=2$ and $N=2$.

7. EMPIRICAL EXAMPLES

We investigate structural changes in the conditional mean and in the error variance of US inflation, quarterly from 1959:1 to 2018:4. For comparison purposes, we use Stock and Watson's (2002) transformation to achieve stationarity, that is, we transform the GDP deflator (X_t) into annual changes of the quarterly inflation rate as $Y_t = 100[\ln(X_t/X_{t-1}) - \ln(X_{t-4}/X_{t-5})]$. The resulting series is presented in Figure 1. We use a simple AR(4) model of the form $Y_t=\mu+\sum_{j=1}^4 \phi_j Y_{t-j}+e_t$. Using the sample from 1959:1 to 2002:3 and a two-step procedure, Stock and Watson (2002) found strong evidence of a structural change in the conditional mean but no or weak evidence of changes in the error variance. Table 12(a) reports the $\sup \mathrm{LR}_{4,T}$ and the $\mathrm{UD}\max \mathrm{LR}_{4,T}$ tests. They suggest at least one change in either or both the coefficients and the variance. Table 12(b) presents the results when testing for changes in the coefficients, allowing for changes in the variance. As in Stock and Watson (2002), we obtain strong evidence of a change in the conditional mean coefficients if we assume no change in the error variance ($\sup \mathrm{LR}_{3,T}$ with $m_a=1$ and $\mathrm{UD}\max \mathrm{LR}_{3,T}$ tests, both with $n_a=0$). The sequential procedure using the $\sup \mathrm{Seq}_{9,T}$ test confirms that a one break specification is preferred with the break date estimated at 1982:1. However, any evidence of changes in the conditional mean disappears once we jointly consider structural changes in the error variance. To assess whether changes in variance are indeed present when accounting

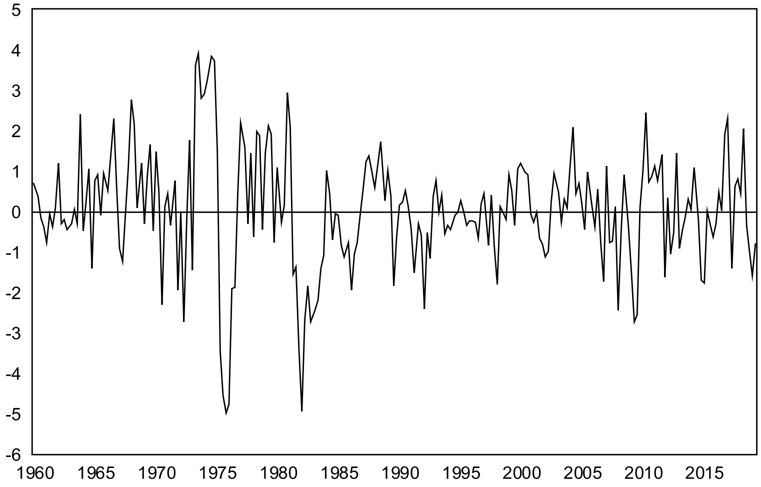

FIGURE 1. Annual change of the quarterly US inflation rate: 1959:1–2018:4.

TABLE 12. Empirical results for the inflation rate.

(a) Tests for structural changes in mean and/or variance

	$\sup \mathrm{LR}_{4,T}$			$\mathrm{UD}\max \mathrm{LR}_{4,T}$
	$m_a=1$	$m_a=2$	$m_a=3$	$M=3, N=3$
$n_a=1$	12.18^2	10.78	9.58	15.91^3
$n_a=2$	15.27^3	13.33^3	11.81^2	
$n_a=3$	15.91^3	15.06^3	14.03^3	

(b) Tests for structural changes in mean

	$\sup \mathrm{LR}_{3,T}$			$\mathrm{UD}\max \mathrm{LR}_{3,T}$	$\sup \mathrm{Seq}_{9,T}$			
	$m_a=1$	$m_a=2$	$m_a=3$	$M=3$	$m_a=1$	$m_a=2$	$m_a=3$	Break Dates
$n_a=0$	22.50^2	19.42^3	15.93^2	22.50^2	10.17	9.38	4.59	1982:1
$n_a=1$	8.54	7.57	7.04	8.54	6.19	6.99	4.59	
$n_a=2$	5.72	6.62	7.37	7.37	2.79	4.96	3.10	
$n_a=3$	9.90	9.72	10.03	10.03	2.74	4.80	4.74	

(c) Tests for structural changes in variance

	$\sup \mathrm{LR}_{2,T}$			$\mathrm{UD}\max \mathrm{LR}_{2,T}$	$\sup \mathrm{Seq}_{10,T}$					
	$n_a=1$	$n_a=2$	$n_a=3$	$N=2$	$n_a=1$	$n_a=2$	$n_a=3$	Break Dates		
$m_a=0$	16.00^3	21.30^3	16.49^3	21.30^3	18.69^3	13.05^2	5.21	1971:3	1983:2	2006:3
$m_a=1$	9.37^2	13.77^3	14.00^3	14.00^3	18.97^3	16.21^3	5.54	1971:3	1982:1	2006:3
$m_a=2$	3.33	8.26^2	11.22^3	11.22^2	18.97^3	16.79^3	6.73			
$m_a=3$	1.69	9.14^2	11.90^3	11.90^2	19.93^3	16.79^3	7.18			

Note: Superscripts 1, 2 and 3 indicate significance at the 10%, 5% and 1% levels, respectively.

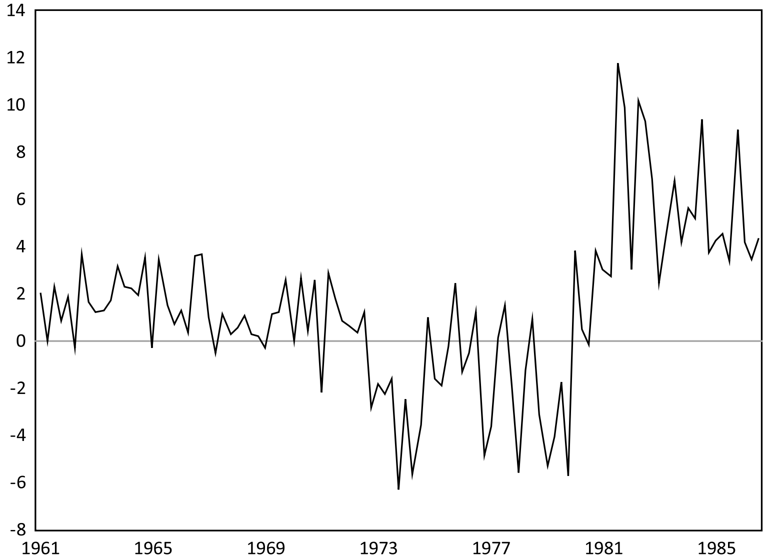

FIGURE 2. US ex post real interest rate: 1961:1–1986:3.

for potential changes in the regression coefficients, Table 12(c) presents the results of the $\sup \mathrm{LR}_{2,T}$ and the $\mathrm{UD}\max \mathrm{LR}_{2,T}$ tests. These suggest the presence of breaks in the variance. The sequential test $\sup \mathrm{Seq}_{10,T}$ suggests 3 breaks at 1971:2, 1983:2, and 2006:3 when $m_a = 0$. Hence, contrary to Stock and Watson (2002), we conclude for 3 structural changes in the error variance and no change in the conditional mean. The changes are such that the variance went from 1.00 to 3.29 in 1971:2, then to 0.49 in 1983:1 and to 1.42 in 2006:3.

We now consider the US ex-post real interest rate and use the same quarterly series from 1961:1–1986:3 (see Figure 2), as in Garcia and Perron (1996) and Bai and Perron (2003a) since it is a widely used example involving important mean shifts, though variance shifts have not been investigated. We use a model with only a constant as regressor (i.e., $z_t = \{1\}$) and account for serial correlations in the errors term via a HAC variance estimator using the hybrid method. The estimate of the scaling factor ψ (see (8)) also uses the hybrid method. Bai and Perron (2003a) found two large mean shifts in 1972:3 and 1980:3 and a small change in 1966:4 using the sequential procedure proposed in Bai and Perron (1998, 2003a), which allows for variance breaks occurring at the same time as the mean breaks, though not at different times. Here, the focus is on assessing whether changes in variance are present and if so whether and how the changes in mean present affect the results. Because they found three breaks in the mean, we use our tests with m_a up to 3 and n_a up to 2. The trimming parameter $\varepsilon = 0.15$ is used. The critical values of both tests when $M = 3$ are provided in Perron and Yamamoto (2019b). Table 13(a) presents the results for the $\sup \mathrm{LR}_{4,T}$ and the $\mathrm{UD}\max \mathrm{LR}_{4,T}$ tests, which suggest clear rejections of the null hypothesis of no breaks. Table 13(b) presents the results when testing for mean breaks accounting for possible variance breaks using the $\sup \mathrm{LR}_{3,T}$ and the $\mathrm{UD}\max \mathrm{LR}_{3,T}$ tests and also the $\sup \mathrm{Seq}_{9,T}$ test to determine the number of breaks. We

TABLE 13. Empirical results for the real interest rate.

(a) Tests for structural changes in mean and/or variance

	$\sup \mathrm{LR}_{4,T}$			$\mathrm{UD}\max \mathrm{LR}_{4,T}$
	$m_a=1$	$m_a=2$	$m_a=3$	$M=3, N=2$
$n_a=1$	8.34^2	4.66	7.50^2	11.44^3
$n_a=2$	8.93^3	11.44^3	6.54^2	

(b) Tests for structural changes in mean

	$\sup \mathrm{LR}_{3,T}$			$\mathrm{UD}\max \mathrm{LR}_{3,T}$	$\sup \mathrm{Seq}_{9,T}$			Break Dates	
	$m_a=1$	$m_a=2$	$m_a=3$	$M=3$	$m_a=1$	$m_a=2$	$m_a=3$		
$n_a=0$	14.66^3	25.75^3	20.60^3	25.75^3	27.86^3	7.63	3.33	1972:3	1980:3
$n_a=1$	8.42^1	25.75^3	24.08^3	25.75^3	25.82^3	6.20	2.99	1972:3	1980:3
$n_a=2$	8.17^1	25.71^3	21.57^3	25.71^3	25.48^3	6.87	3.33	1972:3	1980:3

(c) Tests for structural changes in variance

	$\sup \mathrm{LR}_{2,T}$		$\mathrm{UD}\max \mathrm{LR}_{2,T}$	$\sup \mathrm{Seq}_{10,T}$		Break Dates	
	$n_a=1$	$n_a=2$	$N=2$	$n_a=1$	$n_a=2$		
$m_a=0$	30.03^3	15.96^3	30.03^3	17.05^3	5.89	1972:3	1981:2
$m_a=1$	21.70^3	12.02^3	21.70^3	4.25	6.36	1972:3	
$m_a=2$	16.20^3	10.72^3	16.20^3	15.29^3	6.45	1964:3	1972:3
$m_a=3$	16.42^3	11.62^3	16.42^3	10.88^2	6.45	1966:4	1969:3

Note: Superscripts 1, 2, and 3 indicate significance at the 10%, 5%, and 1% levels, respectively.

obtain evidence for two mean breaks in 1972:3 and 1980:3, irrespective of how many variance breaks are accounted for. However, we do not find evidence for a mean break in 1966:4. To investigate the presence of variance changes, Table 13(c) presents the results of the tests for variance breaks accounting for mean breaks. If we account for no mean breaks ($m_a = 0$), two variance breaks are found in 1972:3 and 1981:2; the former is the same and the latter is close to the dates of the two large mean breaks. However, if one mean break is allowed ($m_a = 1$), only one variance break is found in 1972:3, which suggests that the variance break in 1981:2 was a false rejection due to the ignored mean break. The next issue is whether the 1972:3 variance break is spurious. To see this, we account for two breaks in the mean ($m_a = 2$) and find again two breaks in the variance; one in 1972:3 and the other is in 1964:3. The variance break in 1964:3 is relatively small and was thereby masked when the two mean breaks were not accounted for. More importantly, we again obtain no evidence for a break around 1980:3 but rather one in 1972:3. Therefore, we conclude that both the mean and the variance changed in 1972:3 but only the mean changed in 1980:3, while only the variance changed in 1964:3. This latter change may be responsible for Bai and Perron's (2003a) finding of an additional mean break in 1966:4 using tests that allow for variance changes, though at the same dates as the mean changes. The change are such that the mean went from 1.36 to −1.80 in 1972:3 and to 5.64 in 1980:3, while the variance changed from 1.09 to 1.87 in 1964:3 and then to 6.91 in 1972:3.

8. Conclusion

This paper provided tools for testing for multiple structural breaks in the error variance of a linear regression model with or without the presence of breaks in the regression coefficients. An innovation is that we do not impose any restrictions on the break dates, that is, the breaks in the regression coefficients and in the variance can happen at the same time or at different times. We proposed statistics with asymptotic distributions invariant to nuisance parameters and valid with nonnormal errors and conditional heteroskedasticity, as well as serial correlation. Extensive simulations of the finite sample properties show that our procedures perform well in terms of size and power. A specific to general procedure to estimate the number and type of breaks based on a proposed sequential test is shown to perform well in selecting the number and types of breaks.

Appendix

Proof of Theorem 1. Part (a) follows from Qu and Perron (2007a, Theorem 5) under Assumption A1. For part (b),

$$\begin{aligned}
&\sup \mathrm{LR}_{2,T}(m_a, n_a, \varepsilon | n=0, m_a) \\
&\quad = 2\big[\log \hat{L}_T\big(\tilde{T}_1^c, \dots, \tilde{T}_{m_a}^c; \tilde{T}_1^v, \dots, \tilde{T}_{n_a}^v\big) - \log \tilde{L}_T\big(\hat{T}_1^c, \dots, \hat{T}_{m_a}^c\big)\big] \\
&\quad = T \log \tilde{\sigma}^2 - \sum_{i=1}^{n_a+1} \big(\tilde{T}_i^v - \tilde{T}_{i-1}^v\big) \log \hat{\sigma}_i^2 \\
&\quad = \sum_{i=1}^{n_a} \big[\tilde{T}_{i+1}^v \log \tilde{\sigma}_{1,i+1}^2 - \tilde{T}_i^v \log \tilde{\sigma}_{1,i}^2 - \big(\tilde{T}_{i+1}^v - \tilde{T}_i^v\big) \log \hat{\sigma}_{i+1}^2\big] + \tilde{T}_1^v\big(\log \tilde{\sigma}_{1,1}^2 - \log \hat{\sigma}_1^2\big),
\end{aligned}$$

where $\tilde{\sigma}_{1,i}^2 = (\tilde{T}_i^v)^{-1} \sum_{t=1}^{\tilde{T}_i^v} (y_t - x_t'\tilde{\beta} - z_t'\tilde{\delta}_{t,j})^2$ with $\tilde{\delta}_{t,j} = \tilde{\delta}_j$ for $\hat{T}_{j-1}^c < t \leq \hat{T}_j^c$ (also let $\delta_{t,j}^0 = \delta_j^0$ for $T_{j-1}^{c0} < t \leq T_j^{c0}$) $(j = 1, \dots, m_a + 1)$ and $\hat{\sigma}_i^2 = (\tilde{T}_i^v - \tilde{T}_{i-1}^v)^{-1} \sum_{t=\tilde{T}_{i-1}^v+1}^{\tilde{T}_i^v} (y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j})^2$. Applying a Taylor expansion to $\log \tilde{\sigma}_{1,i+1}^2$, $\log \tilde{\sigma}_{1,i}^2$ and $\log \hat{\sigma}_{i+1}^2$ around $\log \sigma_0^2$, we obtain

$$\sup \mathrm{LR}_{2,T}(m_a, n_a, \varepsilon | n=0, m_a) = \sum_{i=1}^{n_a} \big(F_{1,T}^i + F_{2,T}^i\big) + o_p(1),$$

where

$$\begin{aligned}
F_{1,T}^i &= \big(\sigma_0^2\big)^{-1} \big[\tilde{T}_{i+1}^v \tilde{\sigma}_{1,i+1}^2 - \tilde{T}_i^v \tilde{\sigma}_{1,i}^2 - \big(\tilde{T}_{i+1}^v - \tilde{T}_i^v\big) \hat{\sigma}_{i+1}^2\big] \\
&= \big(\sigma_0^2\big)^{-1} \sum_{t=\tilde{T}_i^v+1}^{\tilde{T}_{i+1}^v} \big[\big(y_t - x_t'\tilde{\beta} - z_t'\tilde{\delta}_{t,j}\big)^2 - \big(y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j}\big)^2\big]
\end{aligned}$$

and

$$\begin{aligned}
F_{2,T}^i &= -(1/2)\left[\tilde{T}_{i+1}^v \left(\frac{\tilde{\sigma}_{1,i+1}^2 - \sigma_0^2}{\sigma_0^2}\right)^2 - \tilde{T}_i^v \left(\frac{\tilde{\sigma}_{1,i}^2 - \sigma_0^2}{\sigma_0^2}\right)^2 - \big(\tilde{T}_{i+1}^v - \tilde{T}_i^v\big) \left(\frac{\hat{\sigma}_{i+1}^2 - \sigma_0^2}{\sigma_0^2}\right)^2\right] \\
&= (1/2)(I + II + III). \qquad \text{(A.1)}
\end{aligned}$$

We first show that $F^i_{1,T} = o_p(1)$. We can express $F^i_{1,T}$ as

$$
\begin{aligned}
&(\sigma_0^2)^{-1}\big[\big(U_{i+1} + X_{i+1}(\beta^0 - \tilde{\beta}) \\
&\quad + Z_{i+1}(\delta^0_{t,j} - \tilde{\delta}_{t,j})\big)'\big(U_{i+1} + X_{i+1}(\beta^0 - \tilde{\beta}) + Z_{i+1}(\delta^0_{t,j} - \tilde{\delta}_{t,j})\big) \\
&\quad - \big(U_{i+1} + X_{i+1}(\beta^0 - \hat{\beta}) \\
&\quad + Z_{i+1}(\delta^0_{t,j} - \hat{\delta}_{t,j})\big)'\big(U_{i+1} + X_{i+1}(\beta^0 - \hat{\beta}) + Z_{i+1}(\delta^0_{t,j} - \hat{\delta}_{t,j})\big)\big] \\
&= (\sigma_0^2)^{-1}\big[(\hat{\beta} - \tilde{\beta})'X'_{i+1}X_{i+1}(\hat{\beta} - \tilde{\beta}) + (\hat{\delta}_{t,j} - \tilde{\delta}_{t,j})'Z'_{i+1}Z_{i+1}(\hat{\delta}_{t,j} - \tilde{\delta}_{t,j}) \\
&\quad + (\hat{\beta} - \tilde{\beta})'X'_{i+1}Z_{i+1}(\hat{\delta}_{t,j} - \tilde{\delta}_{t,j}) + 2(\beta - \hat{\beta})'X'_{i+1}X_{i+1}(\hat{\beta} - \tilde{\beta}) \\
&\quad + 2(\delta^0_{t,j} - \hat{\delta}_{t,j})'Z'_{i+1}Z_{i+1}(\hat{\delta}_{t,j} - \tilde{\delta}_{t,j}) + 2(\hat{\beta} - \tilde{\beta})'X'_{i+1}Z_{i+1}(\delta^0_{t,j} - \hat{\delta}_{t,j}) \\
&\quad + 2(\beta - \hat{\beta})'X'_{i+1}Z_{i+1}(\hat{\delta}_{t,j} - \tilde{\delta}_{t,j}) + 2(\hat{\beta} - \tilde{\beta})'X'_{i+1}U_{i+1} + 2(\hat{\delta}_{t,j} - \tilde{\delta}_{t,j})'Z'_{i+1}U_{i+1}\big].
\end{aligned}
$$

The result follows using the facts that $X'_{i+1}X_{i+1} = O_p(T)$, $Z'_{i+1}Z_{i+1} = O_p(T)$, $X'_{i+1}Z_{i+1} = O_p(T)$, $X'_{i+1}U_{i+1} = O_p(T^{1/2})$ and $Z'_{i+1}U_{i+1} = O_p(T^{1/2})$. Also, under H_0 with Assumption A1, the estimates of the break fractions converge to the true break fractions at a fast enough rate so that the estimates of the parameters of the models are consistent and have the same limit distribution as when the break dates are known. We have: $\beta^0 - \hat{\beta} = O_p(T^{-1/2})$, $\delta^0_{t,j} - \hat{\delta}_{t,j} = O_p(T^{-1/2})$, $\hat{\beta} - \tilde{\beta} = o_p(T^{-1/2})$ and $\hat{\delta}_{t,j} - \tilde{\delta}_{t,j} = o_p(T^{-1/2})$. The last two quantities are $o_p(T^{-1/2})$ since $\sqrt{T}(\hat{\beta} - \beta^0)$ and $\sqrt{T}(\tilde{\beta} - \beta^0)$ have the same limit distribution under H_0, and likewise for $\sqrt{T}(\hat{\delta}_{t,j} - \delta^0_{t,j})$ and $\sqrt{T}(\tilde{\delta}_{t,j} - \delta^0_{t,j})$. For $F^i_{2,T}$,

$$
\begin{aligned}
\sqrt{I} &= (\tilde{T}^v_{i+1})^{-1/2}\sum_{t=1}^{\tilde{T}^v_{i+1}}\big[\big\{(y_t - x'_t\tilde{\beta} - z'_t\tilde{\delta}_{t,j})/\sigma_0\big\}^2 - 1\big] \\
&= (\tilde{T}^v_{i+1})^{-1/2}\sum_{t=1}^{\tilde{T}^v_{i+1}}\big[(u_t/\sigma_0)^2 - 1\big] + o_p(1) \\
&\Rightarrow \sqrt{\psi}W(\lambda^v_{i+1})/\sqrt{\lambda^v_{i+1}}
\end{aligned}
$$

by Assumption A1. Similarly, $\sqrt{II} \Rightarrow \sqrt{\psi}W(\lambda^v_i)/\sqrt{\lambda^v_i}$ and

$$
\begin{aligned}
\sqrt{III} &= \big[(\tilde{T}^v_{i+1} - \tilde{T}^v_i)/T\big]^{-1/2}T^{-1/2}\sum_{t=T^v_i+1}^{T^v_{i+1}}\big[(u_t/\sigma_0)^2 - 1\big] + o_p(1) \\
&= \big[(\tilde{T}^v_{i+1} - \tilde{T}^v_i)/T\big]^{-1/2}\left\{T^{-1/2}\sum_{t=1}^{T^v_{i+1}}\big[(u_t/\sigma_0)^2 - 1\big] - T^{-1/2}\sum_{t=1}^{T^v_i}\big[(u_t/\sigma_0)^2 - 1\big]\right\} + o_p(1) \\
&\Rightarrow \sqrt{\psi}\big[W(\lambda^v_{i+1}) - W(\lambda^v_i)\big]/\sqrt{\lambda^v_{i+1} - \lambda^v_i}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
F_{2,T}^{i} &\Rightarrow -(\psi/2)\left[\frac{W^2(\lambda_{i+1}^{v})}{\lambda_{i+1}^{v}}-\frac{W^2(\lambda_{i}^{v})}{\lambda_{i}^{v}}-\frac{\left(W(\lambda_{i+1}^{v})-W(\lambda_{i}^{v})\right)^2}{\lambda_{i+1}^{v}-\lambda_{i}^{v}}\right] \\
&= (\psi/2)\frac{\left(\lambda_{i}^{v}W(\lambda_{i+1}^{v})-\lambda_{i+1}^{v}W(\lambda_{i}^{v})\right)^2}{\lambda_{i+1}^{v}\lambda_{i}^{v}(\lambda_{i+1}^{v}-\lambda_{i}^{v})}.
\end{aligned}
$$

This yields

$$
\begin{aligned}
\sup \mathrm{LR}_{2,T}(m_a, n_a, \varepsilon | n=0, m_a) &\Rightarrow \sup_{(\lambda_1^v,\dots,\lambda_{na}^v)\in \Lambda_{v,\varepsilon}^{c}} \sum_{i=1}^{n_a}\frac{\psi}{2}\frac{\left(\lambda_{i}^{v}W(\lambda_{i+1}^{v})-\lambda_{i+1}^{v}W(\lambda_{i}^{v})\right)^2}{\lambda_{i+1}^{v}\lambda_{i}^{v}(\lambda_{i+1}^{v}-\lambda_{i}^{v})} \\
&\le \sup_{(\lambda_1^v,\dots,\lambda_{na}^v)\in \Lambda_{v,\varepsilon}} \sum_{i=1}^{n_a}\frac{\psi}{2}\frac{\left(\lambda_{i}^{v}W(\lambda_{i+1}^{v})-\lambda_{i+1}^{v}W(\lambda_{i}^{v})\right)^2}{\lambda_{i+1}^{v}\lambda_{i}^{v}(\lambda_{i+1}^{v}-\lambda_{i}^{v})}
\end{aligned}
$$

because $\Lambda_{v,\varepsilon}^{c} \subseteq \Lambda_{v,\varepsilon}$. For part (c),

$$
\begin{aligned}
&\sup \mathrm{LR}_{3,T}(m_a, n_a, \varepsilon | m=0, n_a) \\
&\quad = 2\left[\log \hat{L}_T(\tilde{T}_1^c,\dots,\tilde{T}_{m_a}^c;\tilde{T}_1^v,\dots,\tilde{T}_{n_a}^v)-\log \tilde{L}_T(\hat{T}_1^v,\dots,\hat{T}_{n_a}^v)\right] \\
&\quad = \sum_{i=1}^{n_a+1}(\hat{T}_i^v-\hat{T}_{i-1}^v)\log\tilde{\sigma}_i^2-\sum_{i=1}^{n_a+1}(\tilde{T}_i^v-\tilde{T}_{i-1}^v)\log\hat{\sigma}_i^2,
\end{aligned}
$$

where $\tilde{\sigma}_i^2 = (\hat{T}_i^v - \hat{T}_{i-1}^v)^{-1}\sum_{t=\hat{T}_{i-1}^v+1}^{\hat{T}_i^v}(y_t - x_t'\tilde{\beta} - z_t'\tilde{\delta})^2$ and $\hat{\sigma}_i^2 = (\tilde{T}_i^v - \tilde{T}_{i-1}^v)^{-1} \times \sum_{t=\tilde{T}_{i-1}^v+1}^{\tilde{T}_i^v}(y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j})^2$. Applying a Taylor expansion on $\log\tilde{\sigma}_i^2$ and $\log\hat{\sigma}_i^2$ around $\log\sigma_{i0}^2$, we obtain

$$
\sup \mathrm{LR}_{3,T}(m_a, n_a, \varepsilon | m=0, n_a) = \sum_{i=1}^{n_a+1}\left(F_{1,T}^i + F_{2,T}^i\right) + o_p(1),
$$

where $F_{1,T}^i = (\hat{T}_i^v - \hat{T}_{i-1}^v)(\tilde{\sigma}_i^2/\sigma_{i0}^2) - (\tilde{T}_i^v - \tilde{T}_{i-1}^v)(\hat{\sigma}_i^2/\sigma_{i0}^2)$ and

$$
F_{2,T}^i = -(1/2)\left[(\hat{T}_i^v - \hat{T}_{i-1}^v)\left([\tilde{\sigma}_i^2-\sigma_{i0}^2]/\sigma_{i0}^2\right)^2 - (\tilde{T}_i^v - \tilde{T}_{i-1}^v)\left([\hat{\sigma}_i^2-\sigma_{i0}^2]/\sigma_{i0}^2\right)^2\right].
$$

We first show that $F_{2,T}^i = o_p(1)$ as follows. We have:

$$
\begin{aligned}
F_{2,T}^i &= -(1/2)\left[(\hat{T}_i^v - \hat{T}_{i-1}^v)\left(\frac{\tilde{\sigma}_i^2-\sigma_{i0}^2}{\sigma_{i0}^2}\right)^2 - (\tilde{T}_i^v - \tilde{T}_{i-1}^v)\left(\frac{\hat{\sigma}_i^2-\sigma_{i0}^2}{\sigma_{i0}^2}\right)^2\right] \\
&= -(1/2)\left[T^{-1}(\hat{T}_i^v - \hat{T}_{i-1}^v)\left[T^{1/2}\left(\frac{\tilde{\sigma}_i^2-\sigma_{i0}^2}{\sigma_{i0}^2}\right)\right]^2 - T^{-1}(\tilde{T}_i^v - \tilde{T}_{i-1}^v)\left[T^{1/2}\left(\frac{\hat{\sigma}_i^2-\sigma_{i0}^2}{\sigma_{i0}^2}\right)\right]^2\right],
\end{aligned}
$$

where $[(\hat{T}_i^v - \hat{T}_{i-1}^v)/T][\sqrt{T}(\tilde{\sigma}_i^2 - \sigma_{i0}^2)/\sigma_{i0}^2]^2$ and $[(\tilde{T}_i^v - \tilde{T}_{i-1}^v)/T][\sqrt{T}(\hat{\sigma}_i^2 - \sigma_{i0}^2)/\sigma_{i0}^2]^2$ have the same limit distribution under Assumption A3. For $F_{1,T}^i$, let $\sigma_0 = \sigma_{10}$ without loss of generality, then

$$\sum_{i=1}^{n_a+1} F_{1,T}^i = (\sigma_0^2)^{-1} \sum_{i=1}^{n_a+1} [(\hat{T}_i^v - \hat{T}_{i-1}^v)\tilde{\sigma}_i^2 - (\tilde{T}_i^v - \tilde{T}_{i-1}^v)\hat{\sigma}_i^2] + (\sigma_0^2)^{-1} \sum_{i=1}^{n_a+1} ([\sigma_{i0}^2 - \sigma_0^2]/\sigma_{i0}^2)[(\hat{T}_i^v - \hat{T}_{i-1}^v)\tilde{\sigma}_i^2 - (\tilde{T}_i^v - \tilde{T}_{i-1}^v)\hat{\sigma}_i^2].$$

The first term becomes

$$\begin{aligned}
&(\sigma_0^2)^{-1} \sum_{i=1}^{n_a+1} [(\hat{T}_i^v - \hat{T}_{i-1}^v)\tilde{\sigma}_i^2 - (\tilde{T}_i^v - \tilde{T}_{i-1}^v)\hat{\sigma}_i^2] \\
&= (\sigma_0^2)^{-1} \sum_{t=1}^{T} \big[(y_t - x_t'\tilde{\beta} - z_t'\tilde{\delta})^2 - (y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j})^2\big] \\
&= (\sigma_0^2)^{-1} \sum_{j=1}^{m_a} \Bigg[\sum_{t=1}^{\tilde{T}_{j+1}^c}(y_t - x_t'\tilde{\beta} - z_t'\tilde{\delta})^2 - \sum_{t=1}^{\tilde{T}_j^c}(y_t - x_t'\tilde{\beta} - z_t'\tilde{\delta})^2 \\
&\quad - \sum_{t=\tilde{T}_j^c+1}^{\tilde{T}_{j+1}^c} (y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{j+1})^2\Bigg] \\
&\quad + (\sigma_0^2)^{-1} \sum_{t=1}^{\tilde{T}_1^c}(y_t - x_t'\tilde{\beta} - z_t'\tilde{\delta})^2 - (\sigma_0^2)^{-1} \sum_{t=1}^{\tilde{T}_1^c}(y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_1)^2 \\
&= (\sigma_0^2)^{-1} \Bigg\{\sum_{j=1}^{m_a}\big[D^r(1, j+1) - D^r(1, j) - D^u(j+1)\big] + D^r(1,1) - D^u(1)\Bigg\},
\end{aligned} \tag{A.2}$$

where $D^r(1, j) = \sum_{t=1}^{\tilde{T}_j^c}(y_t - x_t'\tilde{\beta} - z_t'\tilde{\delta})^2$ and $D^u(j) = \sum_{t=\tilde{T}_{j-1}^c+1}^{\tilde{T}_j^c}(y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_j)^2$. The second term is $o_p(1)$ by Assumption A3. Using similar derivations as in Qu and Perron (2007b), we obtain

$$\begin{aligned}
&D^r(1, j+1) - D^r(1, j) - D^u(j+1) \\
&= -U_{1:j+1}' Z_{1:j+1}(Z_{1:j+1}' Z_{1:j+1})^{-1} Z_{1:j+1}' U_{1:j+1} + U_{1:j}' Z_{1:j}(Z_{1:j}' Z_{1:j})^{-1} Z_{1:j}' U_{1:j} \\
&\quad + U_{j+1}' Z_{j+1}(Z_{j+1}' Z_{j+1})^{-1} Z_{j+1}' U_{j+1} + o_p(1), \\
&\Rightarrow \frac{\|\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)\|^2}{\lambda_{j+1}^c \lambda_j^c(\lambda_{j+1}^c - \lambda_j^c)}
\end{aligned}$$

by Assumption A2. This yields

$$\sup \mathrm{LR}_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a) \Rightarrow \sup_{(\lambda_1^c, \ldots, \lambda_{m_a}^c) \in \Lambda_{c,\varepsilon}^v} \sum_{j=1}^{m_a} \frac{\left\| \lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c) \right\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)},$$

$$\leq \sup_{(\lambda_1^c, \ldots, \lambda_{m_a}^c) \in \Lambda_{c,\varepsilon}} \sum_{j=1}^{m_a} \frac{\left\| \lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c) \right\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)},$$

because $\Lambda_{c,\varepsilon}^v \subseteq \Lambda_{c,\varepsilon}$. For part (d), we have:

$$\begin{aligned}
&\sup \mathrm{LR}_{4,T}(m_a, n_a, \varepsilon | m = n = 0) \\
&\quad = 2\Big[\sup_{(\lambda_1^c, \ldots, \lambda_{m_a}^c; \lambda_1^v, \ldots, \lambda_{n_a}^v) \in \Lambda_\varepsilon} \log \hat{L}_T(T_1^c, \ldots, T_{m_a}^c; T_1^v, \ldots, T_{n_a}^v) - \log \tilde{L}_T \Big] \\
&\quad = 2\big[\log \hat{L}_T(\tilde{T}_1^c, \ldots, \tilde{T}_{m_a}^c; \tilde{T}_1^v, \ldots, \tilde{T}_{n_a}^v) - \log \tilde{L}_T\big] \\
&\quad = T \log \tilde{\sigma}^2 - \sum_{i=1}^{n_a+1} (\tilde{T}_i^v - \tilde{T}_{i-1}^v) \log \hat{\sigma}_i^2 \\
&\quad = \sum_{i=1}^{n_a} \big[\tilde{T}_{i+1}^v \log \tilde{\sigma}_{1,i+1}^2 - \tilde{T}_i^v \log \tilde{\sigma}_{1,i}^2 - (\tilde{T}_{i+1}^v - \tilde{T}_i^v) \log \hat{\sigma}_{i+1}^2\big] + \tilde{T}_1^v(\log \tilde{\sigma}_{1,1}^2 - \log \hat{\sigma}_1^2),
\end{aligned}$$

where $\tilde{\sigma}_{1,i}^2 = (\tilde{T}_i^v)^{-1} \sum_{t=1}^{\tilde{T}_i^v} (y_t - x_t' \tilde{\beta} - z_t' \tilde{\delta})^2$. Applying a Taylor expansion to $\log \tilde{\sigma}_{1,i+1}^2$, $\log \tilde{\sigma}_{1,i}^2$ and $\log \hat{\sigma}_{i+1}^2$ around $\log \sigma_0^2$, we obtain

$$\sup \mathrm{LR}_{4,T}(m_a, n_a, \varepsilon | m = n = 0) = \sum_{i=1}^{n_a} (F_{1,T}^i + F_{2,T}^i) + o_p(1),$$

where the first term is the same as in (A.2), so that

$$\begin{aligned}
\sum_{i=1}^{n_a} F_{1,T}^i &= \sum_{i=1}^{n_a} (\sigma_0^2)^{-1} \big[\tilde{T}_{i+1}^v \tilde{\sigma}_{1,i+1}^2 - \tilde{T}_i^v \tilde{\sigma}_{1,i}^2 - (\tilde{T}_{i+1}^v - \tilde{T}_i^v) \hat{\sigma}_{i+1}^2\big] + (\sigma_0^2)^{-1} \tilde{T}_1^v (\tilde{\sigma}_{1,1}^2 - \hat{\sigma}_1^2) \\
&= (\sigma_0^2)^{-1} \sum_{t=1}^{T} \big[(y_t - x_t' \tilde{\beta} - z_t' \tilde{\delta})^2 - (y_t - x_t' \hat{\beta} - z_t' \hat{\delta}_{t,j})^2\big] \\
&= (\sigma_0^2)^{-1} \left\{ \sum_{j=1}^{m_a} \big[D^r(1, j+1) - D^r(1, j) - D^u(j+1)\big] + D^r(1,1) - D^u(1) \right\}
\end{aligned}$$

as shown in part (c). The second term is the same as (A.1) but with no changes in δ to construct $\tilde{\sigma}_{1,i}^2$, that is, LR_v defined by (11). Hence,

$$F_{2,T}^i = -(1/2) \left[\tilde{T}_{i+1}^v \left(\frac{\tilde{\sigma}_{1,i+1}^2 - \sigma_0^2}{\sigma_0^2} \right)^2 - \tilde{T}_i^v \left(\frac{\tilde{\sigma}_{1,i}^2 - \sigma_0^2}{\sigma_0^2} \right)^2 - (\tilde{T}_{i+1}^v - \tilde{T}_i^v) \left(\frac{\hat{\sigma}_{i+1}^2 - \sigma_0^2}{\sigma_0^2} \right)^2 \right]$$

as shown in part (b). From the proof of part (c),

$$\sum_{i=1}^{n_a} F_{1,T}^i \Rightarrow \sum_{j=1}^{m_a} \frac{\left\| \lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c) \right\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)}$$

under Assumption A2 and from that of part (b),

$$F_{2,T}^i \Rightarrow \frac{\psi}{2} \frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v)\right)^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)}$$

under Assumption A1. Hence, we obtain

$$\sup \mathrm{LR}_{4,T}(m_a, n_a, \varepsilon | m = n = 0) \Rightarrow \sup_{(\lambda_1^c, \ldots, \lambda_{m_a}^c; \lambda_1^v, \ldots, \lambda_{n_a}^v) \in \Lambda_\varepsilon} \left[\sum_{j=1}^{m_a} \frac{\left\| \lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c) \right\|^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} + \frac{\psi}{2} \sum_{i=1}^{n_a} \frac{\left(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v)\right)^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \right]. \qquad \square$$

References

Andrews, D. W. K. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation." *Econometrica*, 59, 817–858. [1029, 1036]

Andrews, D. W. K. (1993), "Tests for parameter instability and structural change with unknown change point." *Econometrica*, 61, 821–856. [1020, 1021, 1035]

Aue, A., S. Hormann, L. Horváth, and M. Reimherr (2009), "Break detection in the covariance structure of multivariate time series models." *The Annals of Statistics*, 37, 4046–4087. [1020]

Bai, J. (2000), "Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices." *The Annals of Economics and Finance*, 1, 303–339. [1020]

Bai, J., R. L. Lumsdaine, and J. H. Stock (1998), "Testing for and dating breaks in multivariate time series." *Review of Economic Studies*, 65, 395–432. [1020]

Bai, J. and S. Ng (2005), "Tests for skewness, kurtosis, and normality for time series data." *Journal of Business and Economic Statistics*, 23, 49–60. [1036]

Bai, J. and P. Perron (1998), "Estimating and testing linear models with multiple structural changes." *Econometrica*, 66, 47–78. [1020, 1021, 1029, 1032, 1033, 1034, 1047]

Bai, J. and P. Perron (2003a), "Computation and analysis of multiple structural change models." *Journal of Applied Econometrics*, 18, 1–22. [1020, 1047, 1048]

Bai, J. and P. Perron (2003b), "Critical values for multiple structural change tests." *Econometrics Journal*, 6, 72–78. [1029, 1033, 1034]

Bai, J. and P. Perron (2006), "Multiple structural change models: A simulation analysis." In *Econometric Theory and Practice: Frontiers of Analysis and Applied Research* (D. Corbae, S. Durlauf, and B. E. Hansen, eds.), 212–237, Cambridge University Press, Cambridge. [1032, 1044]

Blanchard, O. and J. Simon (2001), "The long and large decline in U.S. output volatility." *Brookings Papers on Economic Activity*, 1, 135–173. [1021]

Brown, R. L., J. Durbin, and J. M. Evans (1975), "Techniques for testing the constancy of regression relationships over time." *Journal of the Royal Statistical Society, Series B*, 37, 149–163. [1020, 1036]

Casini, A. and P. Perron (2019a), "Continuous record asymptotics for structural change models." Unpublished manuscript, Department of Economics, Boston University. [1020]

Casini, A. and P. Perron (2019b), "Structural breaks in time series." In *Oxford Research Encyclopedia of Economics and Finance*. [1019]

Cavaliere, G. and I. Georgiev (2019), "Inference under random limit bootstrap measures." Unpublished manuscript, University of Bologna. [1027]

Davis, R. A., D. Huang, and Y. C. Yao (1995), "Testing for a change in the parameter values and order of an autoregressive model." *The Annals of Statistics*, 23, 282–304. [1020]

Deng, A. and P. Perron (2008), "The limit distribution of the CUSUM of squares test under general mixing conditions." *Econometric Theory*, 24, 809–822. [1020, 1036]

Gadea, M. D., A. Gómez-Loscos, and G. Pérez-Quirós (2018), "Great moderation and great recession: From plain sailing to stormy seas?" *International Economic Review*, 59, 2297–2321. [1021]

Garcia, R. and P. Perron (1996), "An analysis of the real interest rate under regime shifts." *Review of Economics and Statistics*, 78, 111–125. [1047]

Górecki, T., L. Horváth, and P. Kokoszka (2018), "Change point detection in heteroscedastic time series." *Econometrics and Statistics*, 7, 63–88. [1022]

Hansen, B. E. (2000), "Testing for structural change in conditional models." *Journal of Econometrics*, 97, 93–115. [1027]

Hawkins, D. M. (1976), "Point estimation of the parameters of piecewise regression models." *Applied Statistics*, 25, 51–57. [1020]

Herrera, A. M. and E. Pesavento (2005), "The decline in U.S. output volatility: Structural changes and inventory investment." *Journal of Business and Economic Statistics*, 23, 462–472. [1021]

Horváth, L. (1993), "The maximum likelihood method for testing changes in the parameters of normal observations." *The Annals of Statistics*, 21, 671–680. [1020]

Inclán, C. and G. C. Tiao (1994), "Use of cumulative sums of squares for retrospective detection of changes of variance." *Journal of the American Statistical Association*, 89, 913–923. [1020]

Kejriwal, M. (2009), "Tests for a mean shift with good size and monotonic power." *Economics Letters*, 102, 78–82. [1029]

Kejriwal, M. and P. Perron (2008), "The limit distribution of the estimates in cointegrated regression models with multiple structural changes." *Journal of Econometrics*, 146, 59–73. [1020]

Kejriwal, M. and P. Perron (2010), "Testing for multiple structural changes in cointegrated regression models." *Journal of Business and Economic Statistics*, 28, 503–522. [1020, 1029]

Kim, C.-J. and C. R. Nelson (1999), "Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle." *Review of Economics and Statistics*, 81, 608–616. [1021]

Kurozumi, E. and P. Tuvaandorj (2011), "Model selection criteria in multivariate models with multiple structural changes." *Journal of Econometrics*, 164, 218–238. [1020]

Liu, J., S. Wu, and J. V. Zidek (1997), "On segmented multivariate regressions." *Statistica Sinica*, 7, 497–525. [1020]

McConnell, M. M. and G. Pérez-Quirós (2000), "Output fluctuations in the United States: What has changed since the early 1980's?" *American Economic Review*, 90, 1464–1476. [1020, 1021]

Perron, P. (1989), "The great crash, the oil price shock, and the unit root hypothesis." *Econometrica*, 57, 1361–1401. [1022]

Perron, P. (2006), "Dealing with structural breaks." In *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory* (K. Patterson and T. C. Mills, eds.), 278–352, Palgrave Macmillan, Basingstoke. [1019]

Perron, P. and Z. Qu (2006), "Estimating restricted structural change models." *Journal of Econometrics*, 134, 373–399. [1020]

Perron, P. and Y. Yamamoto (2014), "A note on estimating and testing for multiple structural changes in models with endogenous regressors via 2SLS." *Econometric Theory*, 30, 491–507. [1020]

Perron, P. and Y. Yamamoto (2015), "Using OLS to estimate and test for structural changes in models with endogenous regressors." *Journal of Applied Econometrics*, 30, 119–144. [1020]

Perron, P. and Y. Yamamoto (2019a), "Pitfalls of two steps testing for changes in the error variance and coefficients of a linear regression models." *Econometrics*, 7, 22. [1021]

Perron, P. and Y. Yamamoto (2019b), "The great moderation: Updated evidence with joint tests for multiple structural changes in variance and persistence." Unpublished manuscript, Department of Economics, Boston University. [1021, 1034, 1047]

Perron, P., Y. Yamamoto, J. Zhou (2020), "Supplement to 'Testing jointly for structural changes in the error variance and coefficients of a linear regression model'." *Quantitative Economics Supplemental Material*, 11, https://doi.org/10.3982/QE1332. [1022]

Pitarakis, J.-Y. (2004), "Least-squares estimation and tests of breaks in mean and variance under misspecification." *Econometrics Journal*, 7, 32–54. [1021]

Qu, Z. and P. Perron (2007a), "Estimating and testing multiple structural changes in multivariate regressions." *Econometrica*, 75, 459–502. [1020, 1021, 1023, 1025, 1027, 1029, 1049]

Qu, Z. and P. Perron (2007b), "Supplement to 'Estimating and testing multiple structural changes in multivariate regressions'." *Econometrica Supplemental Material*, 75, https://www.econometricsociety.org/sites/default/files/ECTA5733SUPP_0.pdf. [1052]

Sensier, M. and D. van Dijk (2004), "Testing for volatility changes in U.S. macroeconomic time series." *Review of Economics and Statistics*, 86, 833–839. [1021]

Stock, J. H. and M. W. Watson (2002), "Has the business cycle changed and why?" In *NBER Macroeconomics Annual*, Vol. 17 (M. Gertler and K. Rogoff, eds.), 159–218, MIT Press, Cambridge, MA. [1021, 1045, 1047]

Xu, K.-L. (2013), "Powerful test for structural changes in volatility." *Journal of Econometrics*, 173, 126–142. [1020]

Co-editor Tao Zha handled this manuscript.

Manuscript received 16 April, 2019; final version accepted 8 February, 2020; available online 11 February, 2020.