HIAS-E-85

Testing Jointly for Structural Changes in the Error Variance and Coefficients of a Linear Regression Model

Pierre Perron[a], Yohei Yamamoto[b] and Jing Zhou[c]

(a) Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215
(b) Department of Economics, Hitotsubashi University, 2-1 Naka Kunitachi, Tokyo 186-8601 Japan
(c) Seeking Sense (Shanghai) Investment Management Co., Ltd., Taiping Finance Tower, 488 Middle Yincheng Road, Shanghai, China 200120

April, 2019

Hitotsubashi Institute for Advanced Study, Hitotsubashi University
2-1, Naka, Kunitachi, Tokyo 186-8601, Japan
tel:+81 42 580 8604 http://hias.ad.hit-u.ac.jp/

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

Pierre Perron[†] Yohei Yamamoto[‡]

Boston University Hitotsubashi University

Jing Zhou[§]

Seeking Sense Investment Management Co., Ltd.

April 16, 2019

Abstract

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 non-Normal 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 pre-specified 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 has been a concern.

JEL Classification: C22; **Keywords**: Change-point; Variance shift; Conditional heteroskedasticity; Likelihood ratio tests.

---

*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.

[†]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

[‡]Department of Economics, Hitotsubashi University, 2-1 Naka Kunitachi, Tokyo, 186-8601 Japan (yohei.yamamoto@econ.hit.ac.jp).

[§]Seeking Sense (Shanghai) Investment Management Co., Ltd., Taiping Finance Tower, 488 Middle Yincheng Road, Shanghai, China 200120 (zhoujing112@foxmail.com)

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, 2019). The problem of multiple structural changes has received attention mostly in the context of a single regression. Bai and Perron (1998, 2003a) provide a comprehensive treatment of various issues: 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. Perron and Qu (2006) extend this analysis to the case where arbitrary linear restrictions are imposed on the coefficients of the model. Related contributions include Hawkins (1976) who presents a comprehensive treatment of estimation based on a dynamic programming algorithm. Also, Kurozumi and Tuvaandorj (2011) propose an information criterion for the selection of the number of changes; see also Liu, Wu and Zidek (1997). Bai, Lumsdaine and Stock (1998) consider 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) considers 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) deal with issues related to testing and inference with multiple structural changes in a single equation cointegrated model. Perron and Yamamoto (2014) derive 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 (2018) provides 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, i.e., confidence intervals with adequate coverage rates and smaller average lengths, especially for small breaks.

With respect to testing for changes in the variance of the regression error, the results are quite sparse. Horváth (1993) considers 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) extend the analysis to an autoregressive process under similar conditions. Aue et. al. (2009) propose non-parametric tests for changes in the variances or autocovariances of multivariate linear or non-linear time series models. Deng and Perron (2008) extended the CUSUM of squares test of Brown, Durbin and Evans (1975) allowing very general conditions on the regressors and the errors (as suggested by Inclán and Tiao, 1994, for Normally distributed time series). Xu (2013) provide a further extension with a robust estimate of the long-run variance of the squared errors of closer relevance to our objectives, Qu and Perron (2007a) consider 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 the work of Qu and Perron (2007a) to provide a comprehensive treatment of 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, allowing the break dates to be different or overlap. 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 non-Normal 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 pre-specified 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.

These testing problems are important for practical applications; e.g., documenting structural changes in the variability of shocks to simple autoregressions or Vector Autoregressive Models; see Blanchard and Simon (2001), Herrera and Pesavento (2005), Kim and Nelson (1999), McConnell and Perez-Quiros (2000), Sensier and van Dijk (2004) and Stock and Watson (2002). Given the lack of proper testing procedures, a common approach is to apply standard 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; see, e.g., Herrera and Pesavento (2005) and Stock and Watson (2002). This is a rather ad hoc procedure. To test for a change in variance only (imposing no change in the regression coefficients), only can apply a CUSUM of squares test to estimated residuals. This test is, however, adequate only if no change in coefficient is present. It is often the case that changes in both coefficients and variance occur and the break dates need not be the same. 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 coefficient when testing for changes in variance induces both size distortions and a loss of power. See Perron and Yamamoto (2019a) and Pitarakis (2004) for extensive documentation about these issues. Hence, what is needed is a joint approach. To do so, our testing procedures are based on quasi likelihood ratio tests constructed using a likelihood function for identically and independently distributed Normal errors. We then apply corrections to have limit distributions free of nuisance parameters in the presence of non-Normal distribution and conditional heteroskedasticity. We also consider extensions that allow for serial correlation. For applications of the methods proposed, see Gadea et al. (2018) and Perron and Yamamoto (2019b).

The paper is structured as follows. Section 2 presents the class of models and the 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 under the alternative hypothesis, 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 brief concluding remarks and an appendix contains some technical derivations. An online supplement contains additional material.

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:

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

for $j = 1, ..., m+1$. 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 $(j = 1, ..., m+1)$ are the corresponding vectors of coefficients; u_t is the disturbance at time t. The break dates $(T_1^c, ..., T_m^c)$ 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 when all coefficients are subject to change. We also allow for n breaks (or $n+1$ regimes) for the variance of the errors, $E(u_t^2) = \sigma^2$, occurring at unknown dates $(T_1^v, ..., T_n^v)$. Accordingly, the error term u_t has zero mean and variance σ_i^2 for $T_{i-1}^v + 1 \leq t \leq T_i^v$ $(i = 1, ..., n+1)$, 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 $(T_1^c, ..., T_m^c)$ and the n-vector $(T_1^v, ..., T_n^v)$ 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 $Y = (y_1, ..., y_T)'$, $X = (x_1, ..., x_T)'$, $U = (u_1, ..., u_T)'$, $\delta = (\delta_1', ..., \delta_{m+1}')'$, and $\bar{Z}$ diagonally partitions Z at $(T_1^c, ..., T_m^c)$, i.e., $\bar{Z} = diag(Z_1, ..., Z_{m+1})$ with $Z_i = (z_{T_{i-1}^c+1}, ..., z_{T_i^c})'$. The true value of the parameters are $\delta^0 = (\delta_1^{0'}, ..., \delta_{m+1}^{0'})'$ and $(T_1^{c0}, ..., T_m^{c0})$ and $\bar{Z}^0$ diagonally partitions Z at $(T_1^{c0}, ..., T_m^{c0})$. Hence, the data-generating process (DGP) is:

$$Y = X\beta^0 + \bar{Z}^0\delta^0 + U \tag{2}$$

with $E(UU') = \Omega^0$, where the diagonal elements of Ω^0 are σ_{i0}^2 for $T_{i-1}^{v0} + 1 \leq t \leq T_i^{v0}$ $(i = 1, ..., n+1)$. 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). The method of estimation considered is quasi maximum likelihood (QML) assuming serially uncorrelated Gaussian errors. They prove consistency of the estimates of the break fractions $(\lambda_1^0, ..., \lambda_K^0) \equiv (T_1^0/T, ..., T_K^0/T)$, where T_i^0 $(i = 1, ..., K)$ denotes the union of the elements of $(T_1^{c0}, ..., T_m^{c0})$ and $(T_1^{v0}, ..., T_n^{v0})$. This is done under general conditions on the regressors and the errors; see Section 4. The conditions are mild and allow for substantial conditional heteroskedasticity and autocorrelation.

The testing problems to be considered 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 non-negative integers. These are useful to assess the adequacy of a model with a particular number of breaks by looking at whether including one more break 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:

$$\log \widetilde{L}_T = -(T/2)(\log 2\pi + 1) - (T/2)\log \widetilde{\sigma}^2, \tag{3}$$

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

$$\log \hat{L}_T\left(T_1^v, ..., T_n^v\right) = -(T/2)(\log 2\pi + 1) - \textstyle\sum_{i=1}^{n_a+1}[(T_i^v - T_{i-1}^v)/2]\log \hat{\sigma}_i^2, \tag{4}$$

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 $i = 1, ..., n_a + 1$. Hence, the Sup-LR test is

$$\begin{aligned} \sup LR_{1,T}\left(n_a, \varepsilon | m = n = 0\right) &= \sup_{\left(\lambda_1^v, ..., \lambda_{n_a}^v\right) \in \Lambda_{v,\varepsilon}} 2[\log \hat{L}_T\left(T_1^v, ..., T_{n_a}^v\right) - \log \widetilde{L}_T] \\ &= 2[\log \hat{L}_T(\hat{T}_1^v, ..., \hat{T}_{n_a}^v) - \log \widetilde{L}_T] \end{aligned}$$

where the estimates $(\hat{T}_1^v, ..., \hat{T}_{n_a}^v)$ are the QMLE obtained by imposing the restriction that there is no structural change in the coefficients and

$$\Lambda_{v,\varepsilon} = \{\left(\lambda_1^v, ..., \lambda_{n_a}^v\right);\ \left|\lambda_{i+1}^v - \lambda_i^v\right| \geq \varepsilon\ (i = 1, ..., n_a - 1), \lambda_1^v \geq \varepsilon, \lambda_{n_a}^v \leq 1 - \varepsilon\},$$

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 $\{T_1^c, ..., T_{m_a}^c\}$, the likelihood function under H_0 is $\log \widetilde{L}_T(T_1^c, ..., T_{m_a}^c) = -(T/2)(\log 2\pi + 1) - (T/2)\log \widetilde{\sigma}^2$, where $\widetilde{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta}_{t,j})^2$, $\widetilde{\beta} = (X'M_{\bar{Z}}X)^{-1}X'M_{\bar{Z}}Y$ and $\widetilde{\delta}_{t,j} = (Z_j'Z_j)^{-1}Z_j(Y_j - X_j\widetilde{\beta})$ for $T_{j-1}^c < t \leq T_j^c$, with $M_{\bar{Z}} = I - \bar{Z}\left(\bar{Z}'\bar{Z}\right)^{-1}\bar{Z}'$, $\bar{Z} = diag\,(Z_1, ..., Z_{m_a+1})$, and $Z_j = (z_{T_{j-1}^c+1}, ..., z_{T_j^c})'$, $Y_j = (y_{T_{j-1}^c+1}, ..., y_{T_j^c})'$, $X_j = (x_{T_{j-1}^c+1}, ..., x_{T_j^c})'$ for $T_{j-1}^c < t \leq T_j^c$ $(j = 1, ..., m_a + 1)$. The log-likelihood value under H_1 is, for given partitions $\{T_1^c, ..., T_{m_a}^c\}$ and $\{T_1^v, ..., T_{n_a}^v\}$,

$$\log \hat{L}_T\left(T_1^c, ..., T_{m_a}^c; T_1^v, ..., T_{n_a}^v\right) = -(T/2)(\log 2\pi + 1) - \textstyle\sum_{i=1}^{n_a+1}[(T_i^v - T_{i-1}^v)/2]\log \hat{\sigma}_i^2, \quad (5)$$

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

$$\begin{aligned}
&\sup LR_{2,T}\,(m_a, n_a, \varepsilon | n = 0, m_a) \\
=\ & 2[\sup_{\left(\lambda_1^c, ..., \lambda_{m_a}^c; \lambda_1^v, ..., \lambda_{n_a}^v\right)\in\Lambda_\varepsilon} \log \hat{L}_T(T_1^c, ..., T_{m_a}^c; T_1^v, ..., T_{n_a}^v) - \sup_{\left(\lambda_1^c, ..., \lambda_{m_a}^c\right)\in\Lambda_{c,\varepsilon}} \log \widetilde{L}_T(T_1^c, ..., T_{m_a}^c)] \\
=\ & 2[\log \hat{L}_T(\widetilde{T}_1^c, ..., \widetilde{T}_{m_a}^c; \widetilde{T}_1^v, ..., \widetilde{T}_{n_a}^v) - \log \widetilde{L}_T(\hat{T}_1^c, ..., \hat{T}_{m_a}^c)],
\end{aligned}$$

where $\Lambda_{c,\varepsilon} = \{(\lambda_1^c, ..., \lambda_m^c)\,;\, \left|\lambda_{j+1}^c - \lambda_j^c\right| \geq \varepsilon\ (j = 1, ..., m_a - 1), \lambda_1^c \geq \varepsilon, \lambda_{m_a}^c \leq 1 - \varepsilon\}$ and

$$\begin{aligned}
\Lambda_\varepsilon \ =\ & \{(\lambda_1^c, ..., \lambda_m^c, \lambda_1^v, ..., \lambda_n^v)\,;\ \text{for } (\lambda_1, ..., \lambda_K) = (\lambda_1^c, ..., \lambda_m^c) \cup (\lambda_1^v, ..., \lambda_n^v) \\
& |\lambda_{j+1} - \lambda_j| \geq \varepsilon\ (j = 1, ..., K - 1), \lambda_1 \geq \varepsilon, \lambda_K \leq 1 - \varepsilon\}
\end{aligned} \quad (6)$$

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 "ˆ" when obtained separately.

The set Λ_ε which defines the possible values of the break fractions in δ $(\lambda_1^c, ..., \lambda_m^c)$ and in σ^2 $(\lambda_1^v, ..., \lambda_m^v)$ 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) $\lambda_1^c \leq \lambda_1^v - \varepsilon$ (the break in δ is before that in σ^2), b) $\lambda_1^c \geq \lambda_1^v + \varepsilon$ (the break in δ is after that 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 $\lambda_1^c \leq \lambda_1^v - \varepsilon$, 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 $\{T_1^v, ..., T_n^v\}$, the likelihood function is $\log \widetilde{L}_T\left(T_1^v, ..., T_{n_a}^v\right) = -(T/2)\left(\log 2\pi + 1\right) - \sum_{i=1}^{n_a+1}[(T_i^v - T_{i-1}^v/2]\log \widetilde{\sigma}_i^2$, where $\widetilde{\sigma}_i^2 = (T_i^v - T_{i-1}^v)^{-1}\sum_{t=T_{i-1}^v+1}^{T_i^v}(y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta})^2$ for $i = 1, ..., n_a + 1$, with $(\widetilde{\beta}', \widetilde{\delta}')' = (W^{\sigma\prime}W^{\sigma})^{-1}W^{\sigma\prime}Y^{\sigma}$, $W^{\sigma} = (w_1^{\sigma}, ..., w_T^{\sigma})'$ 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 (5). The Sup-LR test is

$$\begin{aligned}
& \sup LR_{3,T}\left(m_a, n_a, \varepsilon | m = 0, n_a\right) \\
= \ & 2[\sup_{\left(\lambda_1^c,...,\lambda_{m_a}^c;\lambda_1^v,...,\lambda_{n_a}^v\right)\in\Lambda_\varepsilon} \log \hat{L}_T(T_1^c, ..., T_{m_a}^c; T_1^v, ..., T_{n_a}^v) - \sup_{\left(\lambda_1^v,...,\lambda_{n_a}^v\right)\in\Lambda_{v,\varepsilon}} \log \widetilde{L}_T(T_1^v, ..., T_{n_a}^v)] \\
= \ & 2[\log \hat{L}_T(\widetilde{T}_1^c, ..., \widetilde{T}_{m_a}^c; \widetilde{T}_1^v, ..., \widetilde{T}_{n_a}^v) - \log \widetilde{L}_T(\hat{T}_1^v, ..., \hat{T}_{n_a}^v)]
\end{aligned}$$

For TP-4, under H_0 we have no break and the log-likelihood function is specified by (3). H_1 specifies m_a breaks in δ and n_a breaks in σ^2 and the log likelihood value is given by (5). Hence, the Sup-LR test is

$$\begin{aligned}
& \sup LR_{4,T}\left(m_a, n_a, \varepsilon | n = m = 0\right) \\
= \ & 2[\sup_{\left(\lambda_1^c,...,\lambda_{m_a}^c;\lambda_1^v,...,\lambda_{n_a}^v\right)\in\Lambda_\varepsilon} \log \hat{L}_T\left(T_1^c, ..., T_{m_a}^c; T_1^v, ..., T_{n_a}^v\right) - \log \widetilde{L}_T] \\
= \ & 2[\log \hat{L}_T(\widetilde{T}_1^c, ..., \widetilde{T}_{m_a}^c; \widetilde{T}_1^v, ..., \widetilde{T}_{n_a}^v) - \log \widetilde{L}_T] \qquad (7)
\end{aligned}$$

4 The limiting distributions of the tests

We now consider the limit distribution of the tests, starting with martingale difference errors in Section 4.1 and considering extensions to serially correlated errors in Section 4.2. Section 4.3 deals with double maximum tests and 4.4 with testing for an additional break. Throughout, "$\rightarrow_p$" 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 $\mathcal{F}_t = \sigma$-$field\ \{..., z_{t-1}, z_t, ..., x_{t-1}, x_t, ..., u_{t-2}, u_{t-1}\}$, $E(u_t^2) = \sigma_0^2$ for all t and $T^{-1/2}\sum_{t=1}^{[Ts]}(u_t^2/\sigma_0^2 - 1) \Rightarrow \psi W(s)$, where $W(s)$ is a Wiener process and $\psi = \lim_{T\to\infty} var(T^{-1/2}\sum_{t=1}^{T}(u_t^2/\sigma_0^2 - 1))$.

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 $\mathcal{F}_t = \sigma$-$field\ \{..., z_{t-1}, z_t, ..., x_{t-1}, x_t, ..., u_{t-2}, u_{t-1}\}$, $T^{-1}\sum_{t=1}^{[Ts]} w_t w_t' \to_p sQ$, 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, 2018). This follows from our basic premise that regimes are defined by changes in some coefficients. The second part of A2 assumes no serial correlation in the errors u_t but this will be relaxed later. Since some testing problems imply a non-zero number of breaks under H_0, i.e. 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 $(i = 1, ..., K)$. However, A6 is replaced by (for $i = 1, ..., K$): $\delta_{j+1}^0 - \delta_j^0 = v_T^\delta \delta_j^*$ and $\sigma_{j+1}^0 - \sigma_j^0 = v_T \sigma_j^*$, where $(\delta_j^*, \sigma_j^*) \neq 0$ 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 difference in the modification to A6 is that we require the changes in the variance of the errors to decrease to zero at a slow enough rate as the sample size 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 distribution of the parameter estimates are the same as when the true break dates are known. The reason to require that the change in variance must decrease as T increases is to ensure that A2 holds when changes in variance are permitted under the null hypothesis, in particular when lagged dependent variables are present. Otherwise the limit distribution of the test under TP-3 will not be invariant to nuisance parameters. This should not be constraining in practice since any rate of decrease (subject to some bound described in A3) is permissible; the rate can be as slow as desired. We will show later via simulations that the exact size of the test is close to the nominal level whether the changes in variance are small or very large. To see why this is needed to ensure that A2 is satisfied, let $z_t u_t^\sigma = z_t u_t / \sigma_{i0}$. Then,

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

where $\sigma_0 = \sigma_{10}$ without loss of generality. The result follows since $[(\sigma_{i0}-\sigma_0)/\sigma_{i0}] = O_p(\upsilon_T)$, $\upsilon_T \to 0$ 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:

$$T^{-1}\sum_{t=1}^{[Ts]} z_t z_t' = T^{-1}\sum_{t=1}^{[Ts]} y_{t-1}^2 = T^{-1}\sigma_0^2 \sum_{t=1}^{[Ts]} y_{t-1}^{\sigma 2} + O_p(\upsilon_T) \xrightarrow{p} sQ, \tag{8}$$

where $Q = \sigma_0^2/(1-\beta^2)$ (see Supplement A). The reason why v_T^δ can remain fixed in the case of changes in δ 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 (those associated with the lagged dependent variables) that occur at the same times, which is allowed The limiting distributions, under the relevant null hypothesis, of the likelihood ratio tests for TP-1 to TP-4 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 A1:*

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

b) For TP-2, under A1 and A3,

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

where

$$\begin{aligned}\Lambda_{v,\varepsilon}^c &= \{(\lambda_1^v,\ldots,\lambda_n^v);\ \textit{for}\ (\lambda_1,\ldots,\lambda_K) = (\lambda_1^{0c},\ldots,\lambda_m^{0c}) \cup (\lambda_1^v,\ldots,\lambda_n^v) \\ &\quad |\lambda_{j+1}-\lambda_j| \geq \varepsilon\ (j=1,\ldots,K-1), \lambda_1 \geq \varepsilon, \lambda_K \leq 1-\varepsilon\}\end{aligned}$$

c) For TP-3, under A2 and A3:

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

where

$$\begin{aligned}
\Lambda_{c,\varepsilon}^v &= \{(\lambda_1^c, ..., \lambda_m^c)\,;\ \textit{for}\ (\lambda_1, ..., \lambda_K) = (\lambda_1^c, ..., \lambda_m^c) \cup (\lambda_1^{0v}, ..., \lambda_n^{0v}) \\
& \quad |\lambda_{j+1} - \lambda_j| \geq \varepsilon\ (j = 1, ..., K-1), \lambda_1 \geq \varepsilon, \lambda_K \leq 1 - \varepsilon\}
\end{aligned}$$

d) For TP-4, under A1 and A2:

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

where

$$\begin{aligned}
\Lambda_{cv,\varepsilon} &= \{\left(\lambda_1^c, ..., \lambda_m^c; \lambda_1^v, ..., \lambda_{n_a}^v\right);\ \left|\lambda_{j+1}^c - \lambda_j^c\right| \geq \varepsilon\ (j = 1, ..., m_a - 1), \lambda_1^c \geq \varepsilon, \lambda_{m_a}^c \leq 1 - \varepsilon \\
& \quad \left|\lambda_{i+1}^v - \lambda_i^v\right| \geq \varepsilon\ (i = 1, ..., n_a - 1), \lambda_1^v \geq \varepsilon, \lambda_{n_a}^v \leq 1 - \varepsilon\}.
\end{aligned}$$

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:

$$\hat{\psi} = T^{-1} \sum_{j=-(T-1)}^{T-1} \omega(j, b_T) \sum_{t=|j|+1}^{T} \hat{\eta}_t \hat{\eta}_{t-j} \tag{9}$$

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 residuals. Here $\omega(j, b_T)$ is a weight function and b_T some bandwidth which can be selected using one of the many alternative ways proposed; see, e.g., Andrews (1991). 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), 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 also Kejriwal and Perron, 2010). Simulations showed that using the residuals under H_1 to select b_T and construct the sample autocovariances of η_t leads to tests with important size distortions. Using the residuals under the null for both leads to conservative and less powerful tests. Using the hybrid method permits to control the exact size in small samples without significant loss of power. In our simulations and empirical applications, we use the Quadratic Spectral kernel and to select b_T we adopt the method suggested by 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, $\psi = 2$ so that no adjustment is necessary, a case that was covered by Qu and Perron (2007a). Since these cases are of less practical relevance, we shall only consider a correction involving $\hat{\psi}$ as defined by (9).

The following corrected statistics then have nuisance parameter free limit distributions:

$$\begin{aligned} \sup LR_{1,T}^* &= (2/\hat{\psi}) \sup LR_{1,T} \Rightarrow \sup_{(\lambda_1^v,\ldots,\lambda_{n_a}^v)\in\Lambda_{v,\epsilon}} \sum_{i=1}^{n_a} \frac{\left(\lambda_i^v W\left(\lambda_{i+1}^v\right) - \lambda_{i+1}^v W\left(\lambda_i^v\right)\right)^2}{\lambda_{i+1}^v \lambda_i^v \left(\lambda_{i+1}^v - \lambda_i^v\right)} \qquad (10) \\ \sup LR_{2,T}^* &= (2/\hat{\psi}) \sup LR_{2,T} \Rightarrow \sup_{(\lambda_1^v,\ldots,\lambda_{n_a}^v)\in\Lambda_{v,\epsilon}^c} \sum_{i=1}^{n_a} \frac{\left(\lambda_i^v W\left(\lambda_{i+1}^v\right) - \lambda_{i+1}^v W\left(\lambda_i^v\right)\right)^2}{\lambda_{i+1}^v \lambda_i^v \left(\lambda_{i+1}^v - \lambda_i^v\right)} \\ &\leq \sup_{(\lambda_1^v,\ldots,\lambda_{n_a}^v)\in\Lambda_{v,\epsilon}} \sum_{i=1}^{n_a} \frac{\left(\lambda_i^v W\left(\lambda_{i+1}^v\right) - \lambda_{i+1}^v W\left(\lambda_i^v\right)\right)^2}{\lambda_{i+1}^v \lambda_i^v \left(\lambda_{i+1}^v - \lambda_i^v\right)}. \end{aligned}$$

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

$$\sup LR_{4,T}^* = \sup LR_{4,T} - [(\hat{\psi} - 2)/\hat{\psi}] LR_v \tag{11}$$

where LR_v is the LR test for 0 versus n_a breaks in variance evaluated at $\{\widetilde{T}_1^v, ..., \widetilde{T}_{n_a}^v\}$ obtained by maximizing the likelihood function jointly allowing for m_a breaks in δ, i.e.,

$$LR_v = 2[\log \hat{L}_T(\widetilde{T}_1^v, ..., \widetilde{T}_{n_a}^v) - \log \widetilde{L}_T], \tag{12}$$

where $\log \hat{L}_T(\cdot)$ and $\log \widetilde{L}_T$ are defined by (4) and (3), respectively. Note that LR_v is not equivalent to $LR_{1,T}(n_a, \varepsilon | m = n = 0)$ which is based on the estimates of the break dates for the changes in variance assuming no break in coefficients. Since $\{\widetilde{T}_1^v/T, ..., \widetilde{T}_{n_a}^v/T\}$ are consistent estimates of the break fractions whether $m_a = 0$ or not, we have:

$$LR_v \Rightarrow (\psi/2) \sup_{(\lambda_1^v, ..., \lambda_{n_a}^v) \in \Lambda_\varepsilon} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)}$$

and, hence,

$$\begin{aligned} \sup LR_{4,T}^* &\Rightarrow \sup_{(\lambda_1^c, ..., \lambda_{m_a}^c; \lambda_1^v, ..., \lambda_{n_a}^v) \in \Lambda_\varepsilon} \left[\begin{array}{c} \sum_{j=1}^{m_a} \frac{||\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)||^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\ + \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \end{array} \right] \\ &\leq \sup_{(\lambda_1^c, ..., \lambda_{m_a}^c; \lambda_1^v, ..., \lambda_{n_a}^v) \in \Lambda_{cv,\varepsilon}} \left[\begin{array}{c} \sum_{j=1}^{m_a} \frac{||\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)||^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\ + \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \end{array} \right] \end{aligned} \tag{13}$$

The limit distribution (13) 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 are mutually independent. The number of replications is 10,000 and $T = 1,000$. For each replication, a sum of the supremum of $\sum_{j=1}^{m_a}(||\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)||^2)/\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)$ with respect to $(\lambda_1^c, ..., \lambda_{m_a}^c)$ 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 $(\lambda_1^v, ..., \lambda_{n_a}^v)$ 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 $LR_{1,T}^*$ and sup $LR_{2,T}^*$ statistics are asymptotically invariant to non-Normal errors, serial correlation and conditional heteroskedasticity so that the limit distribution (10) 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: $\sup_{(\lambda_1^c,...,\lambda_{m_a}^c)\in\Lambda_\varepsilon} W_{3,T}\,(m_a, n_a, \varepsilon|m=0, n_a)$, where

$$W_{3,T}\,(m_a, n_a, \varepsilon|m=0, n_a) = T\hat{\delta}'R'(R\hat{V}(\hat{\delta})R')^{-1}R\hat{\delta} \tag{14}$$

with $\hat{\delta} = (\delta_1', ..., \delta_{m_a+1}')'$ the QMLE of δ under a given partition of the sample, R is the conventional matrix such that $(R\delta)' = (\delta_1' - \delta_2', ..., \delta_{m_a}' - \delta_{m_a+1}')$ and $\hat{V}(\hat{\delta})$ is an estimate of the covariance matrix of $\hat{\delta}$ robust to serial correlation and heteroskedasticity, i.e., a consistent estimate of $V(\hat{\delta}) = \text{plim}_{T\to\infty} T\left(\bar{Z}_\sigma^{*\prime}\bar{Z}_\sigma^*\right)^{-1}\Omega_{\bar{Z}_\sigma^*}\left(\bar{Z}_\sigma^{*\prime}\bar{Z}_\sigma^*\right)^{-1}$, where $\bar{Z}_\sigma^* = M_{X_\sigma}\bar{Z}_\sigma$, $\Omega_{\bar{Z}_\sigma^*} = E(\bar{Z}_\sigma^{*\prime}U_b^*U_b^{*\prime}\bar{Z}_\sigma^*)$, $U_b^* = M_{X_\sigma}U_\sigma$, $M_{X_\sigma} = I_T - X_\sigma(X_\sigma'X_\sigma)^{-1}X_\sigma'$, with $\bar{Z}_\sigma = diag\left(Z_1^\sigma, ..., Z_{m_a+1}^\sigma\right)$, $Z_j^\sigma = (z_{T_{j-1}^c+1}^\sigma, ..., z_{T_j^c}^\sigma)'$, $U_\sigma = (u_1^\sigma, ..., u_T^\sigma)'$, $z_t^\sigma = (z_t/\hat{\sigma}_i)$ and $u_t^\sigma = (u_t/\hat{\sigma}_i)$, for $T_{i-1}^{v0} < t \le T_i^{v0}$ $(i = 1, ..., n_a + 1)$. 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 (5), then plug the estimates into (14) 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 $\sup LR_{4,T}$ is asymptotically equivalent to $\sup LR_{4,T}^* = \sup LR_{3,T} + LR_v$. 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 (14), except that one can use z_t instead of z_t^σ since H_0 specifies no break in variance. The second term LR_v is as defined by (12) with $\hat{\psi}$ constructed by (9).

4.3 Double maximum tests

The tests discussed above need prior information about the specification of H_1, i.e., the number of breaks in δ and in σ^2. In practice, researchers may lack such information, hence the need for TP-5 to TP-8. Bai and Perron (1998) proposed so-called double maximum tests to solve this problem in a model with only breaks in δ. They are tests of no break against an unknown number of breaks given some upper bound. They suggested two versions: the first is an equal-weight version labelled $UD\max$, the second applies weights to the individual tests such that the marginal p-values are equal for all number of changes, denoted $WD\max$. Bai and Perron (2006) showed via simulations that the two versions have similar finite sample properties. Hence, we shall only consider the $UD\max$ test given that it is simpler to construct. The double maximum tests can play a significant role in testing for structural changes and it is arguably the most useful tests 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 non monotonic 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 A1:*

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

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

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

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

$$\begin{aligned} UD\max LR_{3,T} &= \max_{1\leq m_a\leq M} m_a^{-1} \sup LR_{3,T}\left(m_a, n_a, \varepsilon | m = 0, n_a\right) \\ &\Rightarrow \max_{1\leq m_a\leq M} m_a^{-1} \sup_{\left(\lambda_1^c,\ldots,\lambda_{m_a}^c\right)\in\Lambda^v_{c,\epsilon}} \sum_{j=1}^{m_a} \frac{||\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)||^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\ &\leq \max_{1\leq m_a\leq M} m_a^{-1} \sup_{\left(\lambda_1^c,\ldots,\lambda_{m_a}^c\right)\in\Lambda_{c,\epsilon}} \sum_{j=1}^{m_a} \frac{||\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)||^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \end{aligned}$$

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

$$UD\max LR_{4,T} = \max_{1\leq n_a\leq N} \max_{1\leq m_a\leq M} \left(n_a + m_a\right)^{-1} \sup LR^*_{4,T}\left(m_a, n_a, \varepsilon | n = m = 0\right)$$

$$\Rightarrow \max_{1\leq n_a\leq N}\max_{1\leq m_a\leq M}(n_a+m_a)^{-1}\sup_{\left(\lambda_1^c,\dots,\lambda_{m_a}^c;\lambda_1^v,\dots,\lambda_{n_a}^v\right)\in\Lambda_\varepsilon}\left[\begin{array}{c}\sum_{j=1}^{m_a}\frac{||\lambda_j^c W_q(\lambda_{j+1}^c)-\lambda_{j+1}^c W_q(\lambda_j^c)||^2}{\lambda_{j+1}^c\lambda_j^c\left(\lambda_{j+1}^c-\lambda_j^c\right)}\\+\sum_{i=1}^{n_a}\frac{(\lambda_i^v W\left(\lambda_{i+1}^v\right)-\lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v\lambda_i^v\left(\lambda_{i+1}^v-\lambda_i^v\right)}\end{array}\right]$$

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

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

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 $(\widetilde{T}_1^c, ..., \widetilde{T}_m^c; \widetilde{T}_1^v, ..., \widetilde{T}_n^v)$ 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. Following Bai and Perron (1998) and Qu and Perron (2007a), the test is

$$\begin{aligned}\sup Seq_{9,T}\left(m+1, n|m, n\right) &= \max_{1\leq j\leq m+1}\sup_{\tau\in\Lambda_{j,\varepsilon}^c}\hat{L}_T(\widetilde{T}_1^c, ..., \widetilde{T}_{j-1}^c, \tau, \widetilde{T}_j^c, ..., \widetilde{T}_m^c; \widetilde{T}_1^v, ..., \widetilde{T}_n^v)\\ &\quad -\hat{L}_T(\widetilde{T}_1^c, ..., \widetilde{T}_m^c; \widetilde{T}_1^v, ..., \widetilde{T}_n^v)\end{aligned}$$

where $\Lambda_{j,\varepsilon}^c = \{\tau; \widetilde{T}_{j-1}^c + (\widetilde{T}_j^c - \widetilde{T}_{j-1}^c)\varepsilon \leq \tau \leq \widetilde{T}_j^c - (\widetilde{T}_j^c - \widetilde{T}_{j-1}^c)\varepsilon\}$. This amounts to performing $m+1$ tests for a single break in δ for each of the $m+1$ regimes defined by the partition $\{\widetilde{T}_1^c, ..., \widetilde{T}_m^c\}$. Note that there are different scenarios when allowing breaks in δ and in σ^2 to happen at different dates, since $(\widetilde{T}_1^c, ..., \widetilde{T}_m^c)$ and $(\widetilde{T}_1^v, ..., \widetilde{T}_n^v)$ 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 $\widetilde{T}_{j-1}^c$ and $\widetilde{T}_j^c$; 2) otherwise, there is one or more variance breaks between $\widetilde{T}_{j-1}^c$ and $\widetilde{T}_j^c$. In either cases, one can appeal to the results of part (c) of Theorem 1 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, $P(\sup Seq_{9,T}(m+1,n|m,n) \leq x) = G_{q,\varepsilon}(x)^{m+1}$, 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 (14) 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

$$\begin{aligned}\sup Seq_{10,T}(m,n+1|m,n) &= (2/\hat{\psi}) \max_{1 \leq i \leq n+1} \sup_{\tau \in \Lambda^v_{i,\varepsilon}} \hat{L}_T(\widetilde{T}^c_1,...,\widetilde{T}^c_m;\widetilde{T}^v_1,...,\widetilde{T}^v_{i-1},\tau,\widetilde{T}^v_i,...,\widetilde{T}^v_m) \\ &\quad - \hat{L}_T(\widetilde{T}^c_1,...,\widetilde{T}^c_m;\widetilde{T}^v_1,...,\widetilde{T}^v_n)\end{aligned}$$

where $\Lambda^v_{i,\varepsilon} = \{\tau; \widetilde{T}^v_{i-1} + (\widetilde{T}^v_i - \widetilde{T}^v_{i-1})\varepsilon \leq \tau \leq \widetilde{T}^v_i - (\widetilde{T}^v_i - \widetilde{T}^v_{i-1})\varepsilon\}$. 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 non-Normal, serially correlated and conditionally heteroskedastic. One can then use part (b) of Theorem 1 to deduce that, under A1 and A3 applied to each segments under H_0: $\lim_{T\to\infty} P(\sup Seq_{10,T}(m,n+1|m,n) \leq x) = G_{1,\varepsilon}(x)^{n+1}$.

5 Monte Carlo experiments

We present the results of simulation experiments to address the following issues: 1) which version of the correction factor $\hat{\psi}$ has better finite sample properties?; 2) whether applying a correction valid under more general conditions than needed is detrimental to the size and power of the test; 3) the finite sample size and power of the tests proposed.

5.1 The choice of $\hat{\psi}$

To address what specific version of the correction factor to use, we consider the size and power of the $\sup LR^*_{4,T}$ test under the following simple DGP with GARCH(1,1) errors:

$$\begin{aligned} y_t &= \mu_1 + \mu_2 1(t > [.25T]) + e_t,\ e_t = u_t\sqrt{h_t},\ u_t \sim i.i.d.\ N(0,1), \\ h_t &= \tau_1 + \tau_2 1(t > [.75T]) + \gamma e^2_{t-1} + \rho h_{t-1}, \end{aligned}$$

with $h_0 = \tau_1/(1-\gamma-\rho)$ and $\tau_1 = 1$. The sample size is $T = 100$ and $\varepsilon = 0.20$. Under H_0, $\mu_2 = \tau_2 = 0$, while under H_1, one break in mean and one break in variance are allowed

($\mu_1 = 0$ under both H_0 and H_1). We consider four versions for the estimate $\hat{\psi}$ as defined by (9): 1) $\hat{\psi} = 2$, i.e., no dependence in η_t is accounted for (labelled "no correction"), 2) using the residuals under H_1 to construct the bandwidth b_T and to estimate the autocovariances of η_t (labelled "alternative"); 3) using the residuals under H_0 instead (labelled "null"); and, as suggested by Kejriwal (2009), 4) using a hybrid method that constructs the bandwidth b_T using the residuals under H_1 but uses the residuals under H_0 to estimate the autocovariances of η_t (labelled "hybrid"). Here and elsewhere, we use 1,000 replications. The reason to include the "no correction" option is to assess which cases (i.e., which combinations of values for ρ and γ) leads to distortions when serial dependence is not accounted for and how well the various suggested options for corrections improve the size.

The results for the exact size of the test (5% nominal size) are presented in Table 2. The critical values are those of the bound of the limit distribution, hence, a conservative size is expected. The results show that the methods "no correction" and "alternative" exhibit substantial size distortions, that increase with γ and ρ, which indicates the extent of the correlation in the squared residuals. The method "null", on the other hand, shows conservative size distortions as expected. The hybrid method shows less conservative size distortions when γ and ρ are not very large. In what follows, we shall also only consider the case with $\rho = 0.2$ and $\gamma = 0.1$,..., 0.5 since they imply tests that require a correction and using either the "null" and "hybrid" methods yields test with good finite samples sizes. The results for power are presented in Table 3. We only consider the methods "null" and "hybrid" given the high size distortions of the methods "no correction" and "alternative". The results show that substantial power gains can be achieved using the "hybrid" method as opposed to the "null" method, especially if the GARCH effect is pronounced. Hence, we recommend using the "hybrid" method and all results below will be based on it.

5.2 Should we always correct?

We now address the issue of whether it is costly in terms of power to use a correction valid under more general conditions than needed. To that effect we first consider the power of the $\sup LR^*_{4,T}$ test under the following DGP with Normal errors:

$$y_t = \mu_1 + \mu_2 1(t > T_1^c) + e_t, \quad e_t \sim \ i.i.d.\ N(0, 1 + \theta 1(t > T_1^v)),$$

where we set $\mu_1 = 0$ and $\mu_2 = \theta$. We consider three scenarios for the timing of the breaks: a common break in mean and variance at $T_1^c = T_1^v = [.5T]$, and disjoint breaks at $\{T_1^c = [.3T]$, $T_1^v = [.6T]\}$ and $\{T_1^c = [.6T]$, $T_1^v = [.3T]\}$. We use $T = 100$, 200 and the power, for

5% nominal size tests, is evaluated at values of θ ranging from 0.25 to 1.5 with $\varepsilon = 0.15$. Three versions of the $\sup LR^{*}_{4,T}$ tests are evaluated: 1) with a full correction based on $\hat{\psi}$ as defined by (9) using the hybrid method (labelled "full"); 2) a correction valid only for *i.i.d.* errors, though not necessarily Normal, given by $\hat{\psi} = \hat{\mu}_4/\hat{\sigma}^4 - 1$, 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 H_0 (labelled "i.i.d."); 3) no correction, i.e., using $\hat{\psi} = 2$, which is the appropriate value with Normal errors (labelled "NC"). The results are presented in Table 4. They show that the power is basically the same using any of the three methods. Hence, there is no cost in using a full correction and we use it throughout.

5.3 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 $\sup LR^{*}_{1,T}(n_a, \varepsilon|m = n = 0)$, abbreviated $\sup LR^{*}_{1,T}(n_a, \varepsilon)$ and the $UD\max LR_{1,T}$ 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

$$CUSQ^{*} = \sup_{\lambda\in[0,1]} |T^{-1/2}[\textstyle\sum_{t=1}^{[T\lambda]}\widetilde{v}_t^2 - ([T\lambda]/T)\sum_{t=1}^{T}\widetilde{v}_t^2]|/\hat{\varphi}_a^{1/2}$$

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 $\hat{\eta}_t = \widetilde{v}_t^2 - \hat{\sigma}$, $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\widetilde{v}_t^2$ and $\widetilde{v}_t$ 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 $\sup LR^{*}_{1,T}(n_a, \varepsilon)$ and $UD\max LR_{1,T}$ tests; b) the relative power of the three tests; c) the power losses obtained when under-specifying the number of breaks; d) the relative power of the $UD\max LR_{1,T}$ compared to the $\sup LR^{*}_{1,T}(n_a, \varepsilon)$ 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

$$\begin{aligned} y_t &= c + \alpha y_{t-1} + e_t,\ e_t = u_t\sqrt{h_t},\ u_t \sim i.i.d.\ N(0,1), \\ h_t &= \tau_1 + \tau_2 1(t > [.5T]) + \gamma e_{t-1}^2 + \rho h_{t-1}, \end{aligned}$$

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 5. The $\sup LR^{*}_{1,T}(1, \varepsilon)$ and $UD\max LR_{1,T}$ 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 $UD\max LR_{1,T}$ test has power close to that of $\sup LR^{*}_{1,T}(1,\varepsilon)$, despite having a broader range of alternatives. The power of the latter two tests dominates that of $CUSQ^{*}$ especially when $T = 100$. Supplement B shows the results to be robust for a static mean model with normal errors.

We now turn to a case with two breaks in variance. The DGP is $y_t = e_t$; $e_t \sim i.i.d.$ $N(0, 1 + \theta 1(T_1^v < t \leq T_2^v))$, i.e., the variance increases at T_1^v and returns to its original level at T_2^v. We consider two scenarios: $\{T_1^v = [.3T],\, T_2^v = [.6T]\}$ and $\{T_1^v = [.2T],\, T_2^v = [.8T]\}$. 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 $UD\max LR_{1,T}$ test with $N = 5$ but include both the $\sup LR^{*}_{1,T}(1,\varepsilon)$ test for a single break and the $\sup LR^{*}_{1,T}(2,\varepsilon)$ test for two breaks to assess the extent of power gains when specifying the correct number of breaks. The results are presented in Table 6. Consider first the size of the tests. The $\sup LR^{*}_{1,T}(1,\varepsilon)$, $\sup LR^{*}_{1,T}(2,\varepsilon)$ and $UD\max LR_{1,T}$ 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 $\sup LR^{*}_{1,T}(1,\varepsilon)$ and $\sup LR^{*}_{1,T}(2,\varepsilon)$ tests, the latter is more powerful, indicating that allowing for the correct number of breaks improves power. The $UD\max LR_{1,T}$ has power between those of the $\sup LR^{*}_{1,T}(1,\varepsilon)$ and $\sup LR^{*}_{1,T}(2,\varepsilon)$ tests. These tests are considerably more powerful than the $CUSQ^{*}$, which has little power.

5.4 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 $\sup LR^{*}_{2,T}(m_a, n_a, \varepsilon | n = 0, m_a)$ 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 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 7. 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 8. It increases rapidly with the magnitude of variance break θ and with the sample size T. It also marginally increases with the value of the trimming ε.

We next investigate the size and power of $\sup LR^{*}_{3,T}(m_a, n_a, \varepsilon | m = 0, n_a)$ 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 errors having a change in variance at $[0.5T]$ with magnitude θ. This is done to investigate potential size distortions due 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 9. 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 10 for the mean model with a coefficient change at $[0.25T]$. The power quickly increases as the break magnitude θ and the sample size T. The power again marginally increases with ε.

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

We now present results about the properties of the $\sup LR^*_{4,T}$ and $UD\max$ 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 $u_t \sim i.i.d.\ N(0,1)$, $h_t = \tau_1 + \gamma e^2_{t-1} + \rho h_{t-1}$, $h_0 = \tau_1/(1-\gamma-\rho)$, $\tau_1 = 1$, $\rho = 0.2$ and γ 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 $\sup LR^*_{4,T}$ 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 are presented in Table 11 and they 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 C shows that the tests have good sizes when the errors are i.i.d. Normal.

We now consider the power of these tests. Since some partial results for the one break case are available in Tables 3-4 for the $\sup LR^*_{4,T}$ 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 12 presents the results for the case with one break in coefficient and two breaks in variance, in which case the DGP is

$$y_t = \mu_1 + \mu_2 1(t > T^c) + e_t;\ e_t \sim i.i.d.\ N(0, 1 + \theta 1(T^v_1 < t \leq T^v_2))$$

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 $\sup LR^*_{4,T}$ 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 variance and the power of the $UD\max$ test is also similar to the power of both versions of the $\sup LR^*_{4,T}$ test. This may, however, be DGP specific. Table 13 presents the results for the case with two breaks in coefficient and one break in variance, with the DGP given by

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

with $\mu_1 = 0$ and $\mu_2 = \theta$. Again, we consider two forms of the $\sup LR^*_{4,T}$ 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 13 shows that for given values of θ and T, the power is lower than for the case of 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, in line with the results of Bai and Perron (2006), the power of the $UD\max$ test is close that attainable when correctly specifying the type and number of breaks.

6 Estimating the numbers of breaks in coefficients and in variance

To select the number of breaks in either the regression coefficient or the error variance, the method suggested is a specific to general procedure that uses the sequential tests proposed in Section 4.4. We determine the number of coefficient breaks and the number of variance breaks, respectively, allowing for a given number of breaks in the other component. More precisely, when selecting the number of breaks in δ, we consider TP-9 and the test $\sup Seq_{9,T}(m+1, n|m, N)$ is applied, starting with $H_0 : m = 0$ and $H_1 : m = 1$, where N is some pre-specified 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 $\sup Seq_{10,T}(m, n+1|M, n)$ is used for $n = 0, 1, ...,$ until a non-rejection occurs. Again, some maximum number of breaks in the coefficients M is imposed.

To assess the finite sample properties of these procedures, we performed a simple simulation experiment. Again, we set $T = 200$ and $\varepsilon = 0.15$ and the basic DGP is:

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

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 also 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 ($T^c = [0.5T], T^v = [0.7T]$); f) a change in both mean and variance occurring at different and distant dates ($T^c = [0.25T], T^v = [0.75T]$). Whenever breaks occur, different magnitudes are considered. The procedure is applied setting the maximum number of breaks to $M = 2$ and $N = 2$, which implies a maximum of four breaks overall, enough for most applications. We also considered a split-sample method discussed in Supplement D.

The results are presented in Tables 14 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 making the correct selection is quite low when the split-sample method is used, e.g., when both changes in mean and variance are not large and occur at different dates, especially when they are far apart. The specific to general approach, however, tests for breaks in the coefficients and in the variance separately allowing for the other components to have unknown breaks, which can avoid segmentations by the other type of breaks and lead to power gains. As shown in Table 14, the probabilities of selecting the correct number of each type of breaks are high with this approach (and higher than with the split-sample method, see Table S.4) when the changes are not large and the break dates are different. Hence, we recommend this procedure in practice.

7 Conclusion

This paper provided tools for testing for multiple structural breaks in the error variance in the 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, i.e., 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 non-Normal 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.

References

Andrews, D.W.K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817-858.

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

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

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

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

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

Bai, J., and Perron, P. (2003a): "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics*, 18, 1-22.

Bai, J., and Perron, P. (2003b): "Critical values for multiple structural change tests," *Econometrics Journal*, 6, 72-78.

Bai, J., and Perron, P. (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.), Cambridge University Press, 2006, 212-237.

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

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

Casini, A., and Perron, P. (2018): "Continuous record asymptotics for structural change models," Unpublished manuscript, Department of Economics, Boston University.

Casini, A., and Perron, P. (2019): "Structural breaks in time series," forthcoming in the *Oxford Research Encyclopedia of Economics and Finance.*

Cavaliere, G., Georgiev, I. (2018): "Inference under random limit bootstrap measures," Unpublished manuscript, University of Bologna.

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

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

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

Goldfeld, S.M., and Quandt, R.E. (1978): "Asymptotic tests for the constancy of regressions in the heteroskedastic case," Econometric Research Program Research Memorandum No. 229, Princeton University.

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

Hawkins, D.M. (1976): "Point estimation of the parameters of piecewise regression models," *Applied Statistics*, 25, 51-57.

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

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

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

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

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

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

Kim, C.-J., and Nelson, C.R. (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.

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

Liu, J., Wu, S., and Zidek, J.V. (1997): "On segmented multivariate regressions," *Statistica Sinica*, 7, 497-525.

McConnell, M.M., and Perez-Quiros, G. (2000): "Output fluctuations in the United States: What has changed since the early 1980's?" *American Economic Review*, 90, 1464-1476.

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

Perron, P., and Qu, Z. (2006): "Estimating restricted structural change models," *Journal of Econometrics* 134, 373-399.

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

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

Perron, P., and Yamamoto, Y. (2019a): "Pitfalls of two steps testing for changes in the error variance and coefficients of a linear regression models," Unpublished manuscript, Department of Economics, Boston University.

Perron, P., and Yamamoto, Y. (2019b): "The great moderation: updated evidence from testing jointly for multiple changes in variance and mean," Unpublished manuscript, Department of Economics, Boston University.

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

Qu, Z., and Perron, P. (2007a): "Estimating and testing multiple structural changes in multivariate regressions," *Econometrica,* 75, 459-502.

Qu, Z., and Perron, P. (2007b): "Supplemental material to: Estimating and testing multiple structural changes in multivariate regressions," (https://www.econometricsociety.org/sites/default/files/ECTA5733SUPP_0.pdf).

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

Stock, J.H., and Watson, M.W. (2002): "Has the business cycle changed and why?" in *NBER Macroeconomics Annual* 17, M. Gertler & K. Rogoff (eds.), MIT press, 159-218.

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

Appendix

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

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

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

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

where

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

and

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

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

$$
(\sigma_0^2)^{-1} \left[\begin{array}{c}
(U_{i+1} + X_{i+1}(\beta^0 - \widetilde{\beta}) \\
+ Z_{i+1}(\delta_{t,j}^0 - \widetilde{\delta}_{t,j}))'(U_{i+1} + X_{i+1}(\beta^0 - \widetilde{\beta}) + Z_{i+1}(\delta_{t,j}^0 - \widetilde{\delta}_{t,j})) \\
-(U_{i+1} + X_{i+1}(\beta^0 - \hat{\beta}) \\
+ Z_{i+1}(\delta_{t,j}^0 - \hat{\delta}_{t,j}))'(U_{i+1} + X_{i+1}(\beta^0 - \hat{\beta}) + Z_{i+1}(\delta_{t,j}^0 - \hat{\delta}_{t,j}))
\end{array} \right]
$$

$$
= (\sigma_0^2)^{-1} \left[\begin{array}{c} (\hat{\beta}-\widetilde{\beta})' X_{i+1}' X_{i+1} (\hat{\beta}-\widetilde{\beta}) + (\hat{\delta}_{t,j}-\widetilde{\delta}_{t,j})' Z_{i+1}' Z_{i+1} (\hat{\delta}_{t,j}-\widetilde{\delta}_{t,j}) \\ +(\hat{\beta}-\widetilde{\beta})' X_{i+1}' Z_{i+1} (\hat{\delta}_{t,j}-\widetilde{\delta}_{t,j}) + 2(\beta-\hat{\beta})' X_{i+1}' X_{i+1} (\hat{\beta}-\widetilde{\beta}) \\ +2(\delta_{t,j}^0-\hat{\delta}_{t,j})' Z_{i+1}' Z_{i+1} (\hat{\delta}_{t,j}-\widetilde{\delta}_{t,j}) + 2(\hat{\beta}-\widetilde{\beta})' X_{i+1}' Z_{i+1} (\delta_{t,j}^0-\hat{\delta}_{t,j}) \\ +2(\beta-\hat{\beta})' X_{i+1}' Z_{i+1} (\hat{\delta}_{t,j}-\widetilde{\delta}_{t,j}) + 2(\hat{\beta}-\widetilde{\beta})' X_{i+1}' U_{i+1} + 2(\hat{\delta}_{t,j}-\widetilde{\delta}_{t,j})' Z_{i+1}' U_{i+1} \end{array} \right].
$$

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, since under H_0 with A1, the estimates of the break fractions converge to the true break fractions at a fast enough rate so that estimates of the parameters of the models are consistent and have the same limit distribution as when the break dates are known, we have: $\beta^0 - \hat{\beta} = O_p(T^{-1/2})$, $\delta_{t,j}^0 - \hat{\delta}_{t,j} = O_p(T^{-1/2})$, $\hat{\beta}-\widetilde{\beta} = o_p(T^{-1/2})$ and $\hat{\delta}_{t,j}-\widetilde{\delta}_{t,j} = o_p(T^{-1/2})$. The last two quantities are $o_p(T^{-1/2})$ since $\sqrt{T}(\hat{\beta}-\beta^0)$ and $\sqrt{T}(\widetilde{\beta}-\beta^0)$ have the same limit distribution under H_0, and likewise for $\sqrt{T}(\hat{\delta}_{t,j}-\delta_{t,j}^0)$ and $\sqrt{T}(\widetilde{\delta}_{t,j}-\delta_{t,j}^0)$. For $F_{2,T}^i$,

$$
\begin{aligned}
\sqrt{I} &= (\widetilde{T}_{i+1}^v)^{-1/2} \sum_{t=1}^{\widetilde{T}_{i+1}^v} \left[(y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta}_{t,j})/\sigma_0]^2 - 1 \right] = (\widetilde{T}_{i+1}^v)^{-1/2} \sum_{t=1}^{\widetilde{T}_{i+1}^v} \left[(u_t/\sigma_0)^2 - 1 \right] + o_p(1) \\
&\Rightarrow \sqrt{\psi} W(\lambda_{i+1}^v)/\sqrt{\lambda_{i+1}^v}
\end{aligned}
$$

by A1. Similarly, $\sqrt{II} \Rightarrow \sqrt{\psi} W(\lambda_i^v)/\sqrt{\lambda_i^v}$ and

$$
\begin{aligned}
\sqrt{III} &= [(\widetilde{T}_{i+1}^v - \widetilde{T}_i^v)/T]^{-1/2} T^{-1/2} \sum_{t=T_i^v+1}^{T_{i+1}^v} [(u_t/\sigma_0)^2 - 1] + o_p(1) \\
&= [(\widetilde{T}_{i+1}^v - \widetilde{T}_i^v)/T]^{-1/2} \left\{ T^{-1/2} \sum_{t=1}^{T_{i+1}^v} [(u_t/\sigma_0)^2 - 1] - \frac{1}{\sqrt{T}} \sum_{t=1}^{T_i^v} [(u_t/\sigma_0)^2 - 1] \right\} + o_p(1) \\
&\Rightarrow \sqrt{\psi} [W(\lambda_{i+1}^v) - W(\lambda_i^v)]/\sqrt{\lambda_{i+1}^v - \lambda_i^v}.
\end{aligned}
$$

Therefore,

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

This yields

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}
\sum_{i=1}^{n_a+1} F_{1,T}^i &= (\sigma_0^2)^{-1} \sum_{i=1}^{n_a+1} \left[(\hat{T}_i^v - \hat{T}_{i-1}^v)\widetilde{\sigma}_i^2 - (\widetilde{T}_i^v - \widetilde{T}_{i-1}^v)\hat{\sigma}_i^2\right] \\
&\quad + (\sigma_0^2)^{-1} \sum_{i=1}^{n_a+1} ([\sigma_{i0}^2 - \sigma_0^2]/\sigma_{i0}^2)\left[(\hat{T}_i^v - \hat{T}_{i-1}^v)\widetilde{\sigma}_i^2 - (\widetilde{T}_i^v - \widetilde{T}_{i-1}^v)\hat{\sigma}_i^2\right].
\end{aligned}$$

The first term becomes,

$$\begin{aligned}
& (\sigma_0^2)^{-1} \sum_{i=1}^{n_a+1} \left[(\hat{T}_i^v - \hat{T}_{i-1}^v)\widetilde{\sigma}_i^2 - (\widetilde{T}_i^v - \widetilde{T}_{i-1}^v)\hat{\sigma}_i^2\right] \\
= \; & (\sigma_0^2)^{-1} \sum_{t=1}^{T} [(y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta})^2 - (y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j})^2], \qquad \text{(A.2)}
\end{aligned}$$

$$
\begin{aligned}
&= (\sigma_0^2)^{-1} \sum_{j=1}^{m_a} \sum_{t=1}^{\hat{T}_{j+1}^c} (y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta})^2 - \sum_{t=1}^{\hat{T}_j^c} (y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta})^2 - \sum_{t=\hat{T}_j^c+1}^{\hat{T}_{j+1}^c} (y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{j+1})^2 \\
&\quad + (\sigma_0^2)^{-1} \sum_{t=1}^{\hat{T}_1^c} (y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta})^2 - (\sigma_0^2)^{-1} \sum_{t=1}^{\hat{T}_1^c} (y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_1)^2, \\
&= (\sigma_0^2)^{-1} \sum_{j=1}^{m_a} \left[D^r(1, j+1) - D^r(1, j) - D^u(j+1) \right] + D^r(1,1) - D^u(1),
\end{aligned}
$$

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

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

by A2. This yields

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

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

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

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

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

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

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

as shown in part (c). The second term is the same as (A.1) but with no changes in δ to construct $\widetilde{\sigma}_{1,i}^2$, i.e., LR_v defined by (12). Hence,

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

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

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

under A2 and from that of part (b),

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

under A1. Hence, we obtain

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

Table 1: Asympotic critical values of the upper bound of the sup $LR^*_{4,T}$ test
(the entries are quantiles x such that $P\left((n_a+m)^{-1}\sup LR^*_4 \leq x\right) \geq \alpha$)

		$\varepsilon=0.10$				$\varepsilon=0.15$				$\varepsilon=0.20$			$\varepsilon=0.25$	$UD\max LR^*_4$		
		$n_a=1$		$n_a=2$		$n_a=1$		$n_a=2$		$n_a=1$		$n_a=2$	$n_a=1$	$M=N=2$		
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	.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
	.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
	.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
	.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	.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
	.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
	.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
	.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	.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
	.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
	.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
	.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	.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
	.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
	.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
	.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	.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
	.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
	.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
	.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

Table 2: Size of the sup $LR^*_{4,T}$ using different estimates of ψ in the case of GARCH(1,1) errors (DGP: $y_t = e_t$, $e_t = u_t\sqrt{h_t}$, with $u_t \sim i.i.d.\ N(0,1)$, $h_t = \tau_1 + \gamma e_{t-1}^2 + \rho h_{t-1}$, $h_0 = \tau_1/(1-\gamma-\rho)$, $\tau_1 = 1$, $T = 100$, $\varepsilon = 0.20$, Alternative hypothesis: $m_a = 1, n_a = 1$).

no correction

$\gamma\backslash\rho$	0.0	0.1	0.2	0.3	0.4	0.5
0.1	0.045	0.049	0.053	0.056	0.064	0.067
0.2	0.087	0.089	0.119	0.113	0.137	0.172
0.3	0.138	0.147	0.171	0.219	0.308	0.354
0.4	0.187	0.249	0.318	0.351	0.431	0.554
0.5	0.280	0.336	0.407	0.479	0.593	-

alternative

$\gamma\backslash\rho$	0.0	0.1	0.2	0.3	0.4	0.5
0.1	0.063	0.066	0.076	0.062	0.079	0.104
0.2	0.076	0.095	0.113	0.111	0.146	0.158
0.3	0.103	0.114	0.147	0.147	0.218	0.279
0.4	0.112	0.139	0.187	0.187	0.289	0.382
0.5	0.142	0.172	0.233	0.233	0.360	-

null

$\gamma\backslash\rho$	0.0	0.1	0.2	0.3	0.4	0.5
0.1	0.032	0.032	0.034	0.027	0.038	0.032
0.2	0.031	0.033	0.037	0.039	0.052	0.052
0.3	0.035	0.022	0.030	0.046	0.045	0.077
0.4	0.025	0.031	0.039	0.047	0.051	0.112
0.5	0.028	0.031	0.036	0.054	0.092	-

hybrid

$\gamma\backslash\rho$	0.0	0.1	0.2	0.3	0.4	0.5
0.1	0.038	0.035	0.037	0.035	0.054	0.055
0.2	0.036	0.041	0.054	0.052	0.059	0.084
0.3	0.040	0.041	0.041	0.063	0.091	0.116
0.4	0.036	0.041	0.062	0.069	0.094	0.166
0.5	0.033	0.047	0.065	0.091	0.122	-

Note: "no correction" specifies $\hat{\psi} = 2$; "alternative" specifies that the unrestricted residuals are used to construct $\hat{\psi}$ and b_T; "null" specifies that the residuals imposing the null hypothesis are used to construct $\hat{\psi}$ and b_T, and "hybrid" specifies that the residuals under the alternative are used to construct b_T and the residuals under the null hypothesis are used to construct $\hat{\psi}$.

Table 3: Power of the sup $LR^*_{4,T}$ using different estimates of ψ in the case of GARCH(1) errors (DGP: $y_t = \mu_1 + \mu_2 1(t > [.25T]) + e_t$, $e_t = u_t\sqrt{h_t}$, with $u_t \sim i.i.d.\ N(0,1)$, $h_t = \tau_1 + \tau_2 1\,(t > [.75T]) + \gamma e_{t-1}^2 + \rho h_{t-1}$, $h_0 = \tau_1/(1-\gamma-\rho)$, $\tau_1 = 1$, $\rho = 0.2$, $T = 100$; $\varepsilon = 0.20$).

a) small change in variance, large change in mean

	$\gamma = 0.1$				$\gamma = 0.3$				$\gamma = 0.5$			
	null		hybrid		null		hybrid		null		hybrid	
$\mu_2\backslash\tau_2$	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5
0.5	0.201	0.222	0.206	0.250	0.148	0.183	0.173	0.197	0.112	0.102	0.147	0.150
1	0.714	0.705	0.719	0.711	0.534	0.565	0.559	0.588	0.399	0.385	0.417	0.391
1.5	0.977	0.979	0.978	0.980	0.911	0.893	0.919	0.901	0.752	0.740	0.760	0.757
2	1.000	1.000	1.000	1.000	0.992	0.997	0.991	0.995	0.944	0.928	0.952	0.923

b) small change in mean, large change in variance

	$\gamma = 0.1$				$\gamma = 0.3$				$\gamma = 0.5$			
	null		hybrid		null		hybrid		null		hybrid	
$\tau_2\backslash\mu_2$	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5
1	0.168	0.287	0.208	0.329	0.115	0.219	0.152	0.259	0.087	0.143	0.130	0.178
3	0.441	0.554	0.609	0.664	0.235	0.359	0.373	0.475	0.159	0.230	0.241	0.295
5	0.586	0.660	0.770	0.843	0.367	0.428	0.548	0.594	0.255	0.286	0.364	0.410
7	0.641	0.732	0.851	0.893	0.453	0.499	0.653	0.664	0.311	0.384	0.445	0.487

Note: "null" specifies that the residuals imposing the null hypothesis are used to construct $\hat{\psi}$ and b_T, and "hybrid" specifies that the residuals under the alternative are used to construct b_T and the residuals under the null hypothesis are used to construct $\hat{\psi}$.

Table 4: Power of the sup $LR^*_{4,T}$ test using different corrections in the case of Normal errors (DGP: $y_t = \mu_1 + \mu_2 1(t > T_1^c) + e_t$; $e_t \sim$ $i.i.d.$ $N(0, 1 + \theta 1(t > T_1^v)$, $\mu_1 = 0$, $\mu_2 = \theta, \varepsilon = 0.15)$

	$T = 100$								
	$T_1^c = T_1^v = [.5T]$			$T_1^c = [.3T]$, $T_1^v = [.6T]$			$T_1^c = [.6T]$, $T_1^v = [.3T]$		
θ	full	i.i.d.	NC	full	i.i.d.	NC	full	i.i.d.	NC
0	0.040	0.032	0.029	0.040	0.032	0.029	0.040	0.032	0.029
0.25	0.120	0.115	0.118	0.108	0.104	0.102	0.106	0.106	0.101
0.5	0.370	0.371	0.370	0.325	0.323	0.328	0.327	0.334	0.330
0.75	0.736	0.727	0.751	0.692	0.689	0.706	0.649	0.647	0.668
1	0.937	0.938	0.941	0.919	0.925	0.936	0.871	0.869	0.877
1.25	0.992	0.992	0.990	0.990	0.990	0.991	0.976	0.980	0.978
1.5	1.000	0.999	0.999	1.000	1.000	1.000	0.991	0.994	0.993
	$T = 200$								
	$T_1^c = T_1^v = [.5T]$			$T_1^c = [.3T]$, $T_1^v = [.6T]$			$T_1^c = [.6T]$, $T_1^v = [.3T]$		
θ	full	i.i.d.	NC	full	i.i.d.	NC	full	i.i.d.	NC
0	0.035	0.036	0.033	0.035	0.036	0.033	0.035	0.036	0.033
0.25	0.227	0.228	0.237	0.168	0.177	0.185	0.199	0.207	0.217
0.5	0.746	0.758	0.764	0.709	0.712	0.712	0.678	0.676	0.673
0.75	0.989	0.987	0.991	0.984	0.982	0.982	0.961	0.963	0.964
1	1.000	0.999	1.000	1.000	1.000	1.000	0.997	0.997	0.998
1.25	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

Note: The nominal size is 5% and 1,000 replications are used. The column "full" refers the test using the correction $\hat{\psi}$ which allows for non-Normal, conditionally heteroskesdatic and serially correlated errors, as defined by (11); the column "i.i.d." refers to a correction that only allows for i.i.d. non-Normal errors, i.e., $\hat{\psi} = \hat{\mu}_4/\hat{\sigma}^4 - 1$, 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 hypotheses; the column "NC" applies no correction and sets $\hat{\psi} = 2$, which is valid with Normal errors.

Table 5: Size and Power of the sup $LR^{*}_{1,T}(n_a = 1, \varepsilon)$, $UD\max$ and $CUSQ^{*}$ tests in a dynamic model with GARCH(1,1) errors (DGP: $y_t = c + \alpha y_{t-1} + e_t$, $e_t = u_t\sqrt{h_t}$, with $u_t \sim i.i.d.\ N(0,1)$, $h_t = \tau_1 + \tau_2 1\,(t > [.5T]) + \gamma e^2_{t-1} + \rho h_{t-1}$, $h_0 = \tau_1 / \,(1 - \gamma - \rho)$, $c = 0.5, \tau_1 = 0.1$, $\rho = 0.2$; $\varepsilon = 0.15$).

	$T = 100$																	
	$\alpha = 0.2$									$\alpha = 0.7$								
	$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$			$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$		
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

Table 6: Size and Power of the sup $LR^*_{1,T}(n_a, \varepsilon)$, $UD\max$ and $CUSQ^*$ tests with Normal errors and two variance breaks (DGP: $y_t = e_t$; $e_t \sim i.i.d.\ N(0, 1 + \theta 1(T_1^v < t \leq T_2^v)$, $T = 200$)

	$T_1^v = [.3T]$, $T_2^v = [.6T]$							$T_1^v = [.2T]$, $T_2^v = [.8T]$						
	$\varepsilon = 0.10$			$\varepsilon = 0.15$				$\varepsilon = 0.10$			$\varepsilon = 0.15$			
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

Table 7: Size of the sup $LR^*_{2,T}(m_a = 1, n_a = 1, \varepsilon | n = 0, m_a = 1)$ test with different trimming parameter ε in the case of Normal Errors
(DGP: $y_t = \mu_1 + \mu_2 1(t > [0.5T]) + e_t$, $e_t \sim i.i.d.\ N(0,1)$, $\mu_1 = 0$).

	$T = 100$				$T = 200$			
$\mu_2 \backslash \varepsilon$	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

Table 8: Power of the sup $LR^*_{2,T}(m_a = 1, n_a = 1, \varepsilon | n = 0, m_a = 1)$ test with different trimming parameter ε in the case of Normal Errors
(DGP: $y_t = \mu_1 + \mu_2 1(t > [0.5T]) + e_t$, $e_t \sim i.i.d.\ N(0, 1 + \theta 1(t > [.25T]))$).

	$T = 100$													
	$\varepsilon = 0.1$							$\varepsilon = 0.2$						
$\theta \backslash \mu_2$	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
	$T = 200$													
	$\varepsilon = 0.1$							$\varepsilon = 0.2$						
$\theta \backslash \mu_2$	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

Table 9: Size of the sup $LR^*_{3,T}(m_a = 1, n_a = 1, \varepsilon | m = 0, n_a = 1)$ test with different trimming parameter ε in the case of Normal Errors

(DGP: $y_t = \mu_1 + e_t$, $e_t \sim i.i.d.\ N(0, 1 + \theta 1(t > [.5T])$, $\mu_1 = 0$).

	$T = 100$				$T = 200$			
$\theta\backslash\varepsilon$	0.1	0.15	0.2	0.25	0.1	0.15	0.2	0.25
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

(DGP: $y_t = c + \alpha y_{t-1} + e_t$, $e_t \sim i.i.d.\ N(0, 1 + \theta 1(t > [.5T])$, $c = 0$, $\alpha = 0.5$).

	$T = 100$				$T = 200$			
$\theta\backslash\varepsilon$	0.1	0.15	0.2	0.25	0.1	0.15	0.2	0.25
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

Table 10: Power of the sup $LR^*_{3,T}(m_a = 1, n_a = 1, \varepsilon | m = 0, n_a = 1)$ test with different trimming parameter ε in the case of Normal Errors

(DGP: $y_t = \mu_1 + \mu_2 1(t > [0.25T]) + e_t$, $e_t \sim i.i.d.\ N(0, 1 + \theta 1(t > [.5T])$, $\mu_1 = 0$).

	$T = 100$													
	$\varepsilon = 0.1$							$\varepsilon = 0.2$						
$\mu_2\backslash\theta$	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
	$T = 200$													
	$\varepsilon = 0.1$							$\varepsilon = 0.2$						
$\mu_2\backslash\theta$	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

Table 11: Size of the sup $LR^*_{4,T}(m_a, n_a)$ and $UD\max$ tests in the case of GARCH(1,1) errors (DGP: $y_t = e_t$, $e_t = u_t\sqrt{h_t}$, with $u_t \sim i.i.d.\ N(0,1)$, $h_t = \tau_1 + \gamma e^2_{t-1} + \rho h_{t-1}$, $\tau_1 = 1$, $\rho = 0.2$, $h_0 = \tau_1/\left(1 - \gamma - \rho\right)$)

	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

Table 12: Power of the sup $LR^*_{4,T}(m_a, n_a)$ and $UD\max$ tests for DGPs with one break in coefficients and two breaks in variance
(DGP: $y_t = \mu_1 + \mu_2 1(t > T^c) + e_t$, $e_t \sim i.i.d.\ N(0, 1 + \theta 1(T_1^v < t \leq T_2^v))$, $\mu_1 = 0, \mu_2 = \theta, \varepsilon = 0.1$)

	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
	T^c= T_1^v= [.3T], T_2^v= [.6T]			T^c= T_2^v = [.6T], T_1^v= [.3T]			T^c= [.3T], T_1^v= [.5T], T_2^v= [.6T]			T^c= [.5T], T_1^v= [.3T], T_2^v= [.6T]			T^c= [.6T], T_1^v= [.3T], T_2^v= [.5T]		
θ	$T = 100$														
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
	$T = 200$														
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	0637	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

Table 13: Power of the sup $LR^*_{4,T}(m_a, n_a)$ and $UD\max$ tests for DGPs with two breaks in coefficients and one break in variance
(DGP: $y_t = \mu_1 + \mu_2 1(T_1^c < t \leq T_2^c) + e_t$, $e_t \sim i.i.d.\ N(0, 1 + \theta 1(t > T^v))$, $\mu_1 = 0, \mu_2 = \theta, \varepsilon = 0.1$).

	m_a= 1 n_a= 1	m_a= 2 n_a= 1	UDmax	m_a= 1 n_a= 1	m_a= 2 n_a= 1	UDmax	m_a= 1 n_a= 1	m_a= 2 n_a= 1	UDmax	m_a= 1 n_a= 1	m_a= 2 n_a= 1	UDmax	m_a= 1 n_a= 1	m_a= 2 n_a= 1	UDmax
	T_1^c= T^v= [.3T], T_2^c= [.6T]			T_1^c= [.3T], T_2^c= T^v= [.6T]			T_1^c= [.5T], T_2^c= [.6T], T^v= [.3T]			T_1^c= [.3T], T_2^c= [.6T], T^v= [.5T]			T_1^c= [.3T], T_2^c= [.5T], T^v= [.6T]		
θ	$T = 100$														
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
	$T = 200$														
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

Table14: Finite sample performance of the specific to general sequential procedure to select the number of breaks in coefficients and variance. (DGP: $y_t = \mu_1 + \mu_2 1(t > T^c) + e_t$, $e_t \sim i.i.d.\ N(0, 1 + \theta 1(t > T^v))$, $\varepsilon = 0.15$, $T = 200$).

	$m=n=0$	$m=n=1$ $T^c=[0.5T], T^v=[0.7T]$				$m=n=1$ $T^c=[0.25T], T^v=[0.75T]$		
		$\mu_2=\theta=1$	$\mu_2=1,\theta=3$	$\mu_2=1,\theta=5$	$\mu_2=\theta=2$	$\mu_2=\theta=1$	$\mu_2=\theta=2$	$\mu_2=1,\theta=3$
$prob(m=0,n=0)$	**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: $prob(m=j,n=i)$ 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 $N=2$.