“Testing jointly for structural changes in the error variance and coefficients of a linear regression model”
by Pierre Perron, Yohei Yamamoto and Jing Zhou
Supplementary Material

A: Proof of the statement in equation (7): Consider an AR(1) model $y_t = \beta y_{t-1} + u_t$ in which the variance of u_t has n breaks and $|\beta| < 1$. Consider the variance adjusted series $y_t^\sigma = \beta y_{t-1}^\sigma + u_t^\sigma$ where $u_t^\sigma = u_t/\sigma_{i0}$ and $y_t^* = \sigma_0 y_t^\sigma$ with $\sigma_0 = \sigma_{10}$, without loss of generality. Then,

$$\begin{aligned} T^{-1}\textstyle\sum_{t=1}^{[Ts]} z_t z_t' &= T^{-1}\textstyle\sum_{t=1}^{[Ts]} y_{t-1}^2 \\ &= T^{-1}\textstyle\sum_{t=1}^{[Ts]} y_{t-1}^{*2} + 2T^{-1}\sum_{t=1}^{[Ts]} y_{t-1}^*(y_{t-1} - y_{t-1}^*) + T^{-1}\sum_{t=1}^{[Ts]}(y_{t-1} - y_{t-1}^*)^2. \end{aligned}$$

We here show that the second and the third terms are $O_p(\upsilon_T)$ and $O_p(\upsilon_T^2)$ where $\upsilon_T \to 0$ uniformly in s. For notational simplicity, we show these results for y_t and y_t^* instead of y_{t-1} and y_{t-1}^* as the difference is negligible. First, for $[T\lambda_{k-1}^{v0}] < t \leq [T\lambda_k^{v0}]$:

$$\begin{aligned} y_t &= \textstyle\sum_{\tau=0}^{t-1} \beta^\tau u_{t-\tau} = \sum_{l=1}^{k} (\sigma_{l,0}/\sigma_0) v_{l,t} \\ y_t^* &= \textstyle\sum_{\tau=0}^{t-1} \beta^\tau u_{t-\tau}^* = \sum_{l=1}^{k} v_{l,t} \end{aligned}$$

where $v_{l,t} = \sigma_0 \sum_{\tau=[T\lambda_{l-1}^{v0}]+1}^{[T\lambda_l^{v0}]} \beta^\tau u_{t-\tau}^\sigma$ for $l < k$ and $v_{k,t} = \sigma_0 \sum_{\tau=[T\lambda_{k-1}^{v0}]+1}^{t} \beta^\tau u_{t-\tau}^\sigma$. These yield

$$\begin{aligned} y_t^*(y_t - y_t^*) &= \textstyle\sum_{l_1=1}^{k}\sum_{l_2=1}^{k} \left(\dfrac{\sigma_{l_2,0} - \sigma_0}{\sigma_0}\right) v_{l_1,t} v_{l_2,t} \\ (y_t - y_t^*)^2 &= \textstyle\sum_{l_1=1}^{k}\sum_{l_2=1}^{k} \left(\dfrac{\sigma_{l_1,0} - \sigma_0}{\sigma_0}\right)\left(\dfrac{\sigma_{l_2,0} - \sigma_0}{\sigma_0}\right) v_{l_1,t} v_{l_2,t} \end{aligned}$$

so that for any $l < k$

$$\begin{aligned} T^{-1}\textstyle\sum_{t=[T\lambda_{l-1}^{v0}]+1}^{[T\lambda_l^{v0}]} y_t^*(y_t - y_t^*) &= \textstyle\sum_{l_1=1}^{k}\sum_{l_2=1}^{k} \underbrace{\left(\dfrac{\sigma_{l_2,0} - \sigma_0}{\sigma_0}\right)}_{=O_p(\upsilon_T)} \underbrace{\left(T^{-1}\textstyle\sum_{t=[T\lambda_{l-1}^{v0}]+1}^{[T\lambda_l^{v0}]} v_{l_1,t} v_{l_2,t}\right)}_{=O_p(1)} \\ &= S_{1,l} = O_p(\upsilon_T) \end{aligned}$$

and

$$\begin{aligned} & T^{-1}\textstyle\sum_{t=[T\lambda_{l-1}^{v0}]+1}^{[T\lambda_l^{v0}]} (y_t - y_t^*)^2 \\ = \; & \textstyle\sum_{l_1=1}^{k}\sum_{l_2=1}^{k} \underbrace{\left(\dfrac{\sigma_{l_1,0} - \sigma_0}{\sigma_0}\right)\left(\dfrac{\sigma_{l_2,0} - \sigma_0}{\sigma_0}\right)}_{=O_p(\upsilon_T^2)} \underbrace{\left(T^{-1}\textstyle\sum_{t=[T\lambda_{l-1}^{v0}]+1}^{[T\lambda_l^{v0}]} v_{l_1,t} v_{l_2,t}\right)}_{=O_p(1)} \\ = \; & S_{2,l} = O_p(\upsilon_T^2). \end{aligned}$$

For any $l < k$, $T^{-1} \sum_{t=[T\lambda_{l-1}^{v0}]+1}^{[T\lambda_l^{v0}]} v_{l_1,t} v_{l_2,t} = O_p(1)$ because $v_{l_1,t}$ and $v_{l_2,t}$ are covariance stationary series for any l_1 and l_2. We can show that the same property holds uniformly in s for $l = k$ with a minor change of notation. Therefore, uniformly in s,

$$\begin{aligned} 2T^{-1} \textstyle\sum_{t=1}^{[Ts]} y_{t-1}^*(y_{t-1} - y_{t-1}^*) &= 2 \textstyle\sum_{l=1}^{k} S_{1,l} = O_p(v_T) \\ T^{-1} \textstyle\sum_{t=1}^{[Ts]} (y_{t-1} - y_{t-1}^*)^2 &= \textstyle\sum_{l=1}^{k} S_{2,l} = O_p(v_T^2) \end{aligned}$$

B: 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:

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

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 (8): 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 S.1. 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. These results dictate our choice of $\rho = 0.2$ and $\gamma = 0.1$,..., 0.5 in the subsequent simulations reported in the text 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 S.2. 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 reported in the main text are based on it.

C: Should we always correct?

We 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 $\hat{\psi}$ 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 S.3. 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 for the results reported in the main text.

D: Local asymptotic power functions.

We consider model (1) focusing on the case of $n = m = 1$ with the following assumptions.
•**Assumption L1:** Assumptions A1 and A3 hold with $\sigma_{20} - \sigma_{10} = \sigma^*/\sqrt{T}$. We also have $T^{-1/2}\sum_{t=1}^{[Ts]}[(u_t^\sigma)^2 - 1] \Rightarrow \psi W(s)$ with $\psi = \lim_{T\to\infty} var(T^{-1/2}\sum_{t=1}^{T}[(u_t^\sigma)^2 - 1])$ and $T^{-1}\sum_{t=1}^{[Ts]}(u_t^\sigma)^2 \xrightarrow{p} s$ uniformly in $s \in [0,1]$.
•**Assumption L2**: Assumptions A2 and A3 hold with $\delta_2^0 - \delta_1^0 = \delta^*/\sqrt{T}$.

In the following, we derive the local asymptotic power of the $\sup LR_{2,T}$ and $\sup LR_{3,T}$

tests, allowing for nuisance breaks, i.e., we consider the tests $\sup LR_{2,T}(n=1,m=1,\varepsilon|n=0,m=1)$ and $\sup LR_{3,T}(m=1,n=1,\varepsilon|m=0,n=1)$. The results are also valid if the nuisance breaks are not accounted for, in which case the tests reduce to the $\sup LR_{1,T}$ and the $\sup LR_T$ test (Andrews, 1993), respectively. Without loss of generality, we denote σ_{20} by σ_0 and δ_2^0 by δ^0.

Lemma S.1 *Under Assumption L1 or Assumption L1 with Assumption A3 allowing for $\delta_1^*=0$,*

$$\sup LR_{2,T}(n=1,m=1,\varepsilon|n=0,m=1) \Rightarrow \sup_{\lambda^v\in\Lambda^c_{v,\epsilon}}(\psi/2)[J(\lambda^v)]^2 \tag{S.1}$$

where

$$\begin{aligned} J(\lambda^v) &= \frac{\lambda^v W(1)-W(\lambda^v)}{\sqrt{\lambda^v(1-\lambda^v)}}+\frac{2\sigma^*\sigma_0}{\sqrt{\psi}}b(\lambda^v) \\ b(\lambda^v) &= \begin{cases} \lambda^{v0}\sqrt{\frac{1-\lambda^v}{\lambda^v}} & \text{if } \lambda^{v0}\le\lambda^v \\ (1-\lambda^{v0})\sqrt{\frac{\lambda^v}{1-\lambda^v}} & \text{if } \lambda^{v0}>\lambda^v \end{cases}. \end{aligned}$$

In particular if $\delta_1^*=0$, which is imposed in the construction of the test,

$$\sup LR_{1,T}(n=1,\varepsilon) \Rightarrow \sup_{\lambda^v\in\Lambda_{v,\varepsilon}}(\psi/2)[J(\lambda^v)]^2. \tag{S.2}$$

Lemma S.2 *Under Assumption L2 or Assumption L2 with Assumption A3 allowing for $\sigma_{10}^*=0$,*

$$\sup LR_{3,T}(m=1,n=1,\varepsilon|m=0,n=1) \Rightarrow \sup_{\lambda^c\in\Lambda^v_{c,\varepsilon}}(\lambda^c)'J(\lambda^c) \tag{S.3}$$

where

$$\begin{aligned} J(\lambda^c) &= \frac{\lambda^c W_q(1)-W_q(\lambda^c)}{\sqrt{\lambda^c(1-\lambda^c)}}+Q^{1/2}\delta^* b(\lambda^c) \\ b(\lambda^c) &= \begin{cases} \lambda^{c0}\sqrt{\frac{1-\lambda^c}{\lambda^c}} & \text{if } \lambda^{c0}\le\lambda^c \\ (1-\lambda^{c0})\sqrt{\frac{\lambda^c}{1-\lambda^c}} & \text{if } \lambda^{c0}>\lambda^c \end{cases} \end{aligned}$$

In particular, if $\sigma_{10}^*=0$, which is imposed in the construction of the test,

$$\sup LR_T(m=1,\varepsilon) \Rightarrow \sup_{\lambda^c\in\Lambda_{c,\varepsilon}} J(\lambda^c)'J(\lambda^c). \tag{S.4}$$

Importantly, the result in (S.4) is the same as Theorem 4(c) in Andrews (1993), if we set (in his notation) $\eta(s)=\delta^* I(s\le\lambda^{c0})$, $S=\sigma_0^2(T^{-1}Z'Z)$ and $M=(T^{-1}Z'Z)$. For comparisons, we also consider the cumulative sum of squares ($CUSQ$) test when there are no nuisance coefficient breaks. With $\hat{u}_t$ the OLS regression residuals, the $CUSQ$ test is:

$$CUSQ = \frac{\max_{1 \le T^v \le T} \left| T^{-1/2} \left[\sum_{t=1}^{T^v} \hat{u}_t^2 - \frac{T^v}{T} \sum_{t=1}^{T} \hat{u}_t^2 \right] \right|}{T^{-1} \sum_{t=1}^{T} \hat{u}_t^2}$$

From Deng and Perron (2008)

$$CUSQ = \sup_{\lambda^v \in [0,1]} \left| T^{-1/2} \left[\sum_{t=1}^{[T\lambda^v]} \left(\frac{u_t^2}{\sigma_0^2} - 1 \right) - \frac{[T\lambda^v]}{T} \sum_{t=1}^{T} \left(\frac{u_t^2}{\sigma_0^2} - 1 \right) \right] \right| + o_p(1)$$

and we obtain the following result.

Lemma S.3 *Under Assumption L1, if there is no structural change in the coefficients, then*

$$CUSQ \Rightarrow \sqrt{\psi} \sup_{\lambda^v \in [0,1]} \left| W(\lambda^v) - \lambda^v W(1) + \frac{2\sigma^* \sigma_0}{\sqrt{\psi}} \mu(\lambda^v) \right| \tag{S.5}$$

where

$$\mu(\lambda^v) = \begin{cases} \lambda^{v0}(1 - \lambda^v) & \text{if } \lambda^{v0} \le \lambda^v \\ \lambda^v(1 - \lambda^{v0}) & \text{if } \lambda^{v0} > \lambda^v \end{cases}.$$

Lemma S.1 suggests that the local asymptotic power of the $\sup LR_{2,T}$ test coincides with that of the $\sup LR_{1,T}$ test except for the fact that the set of permissible break dates $\Lambda_{v,\varepsilon}^c$ becomes smaller than $\Lambda_{v,\varepsilon}$. Lemma S.2 suggests that the local asymptotic power of the $\sup LR_{3,T}$ test coincides with that of the standard $\sup LR_T$ test derived in Theorem 4 of Andrews (1993) except that the set of permissible break dates $\Lambda_{c,\varepsilon}^v$ is in general smaller than $\Lambda_{c,\varepsilon}$. Hence, when testing for changes in variance allowing for changes in coefficients, we have the same local asymptotic power function as when testing for changes in variance when no change in coefficient is present and none is allowed for. Therefore, we incur no loss in local asymptotic power by adopting our more general approach.

We next compare the local asymptotic power functions of the $\sup LR_{1,T}$ test given by (S.2), the $\sup LR_{2,T}$ test given by (S.1) and the $CUSQ$ test given by (S.5) via Monte Carlo simulations. To this end, the Wiener processes $W(\cdot)$ are approximated by the partial sums of i.i.d. standard normal random variables with 5,000 discrete steps. The power functions of 5% nominal size tests are computed based on 10,000 Monte Carlo replications with the value of σ^* ranging from 0 to 10. We also set the trimming $\varepsilon = 0.15$, $\psi = 2$ and $\sigma_0 = 1$, although these particular choices do not qualitatively affect the results. We use the critical values of 8.58 for the $\sup LR_{1,T}$ and $\sup LR_{2,T}$ tests and $\sqrt{2} \times 1.358$ for the $CUSQ$ test.

Figure S.1 shows the asymptotic local power functions of the $\sup LR_{1,T}$ test and the $CUSQ$ test when a break in variance occurs at $\lambda^{v0} = 0.3$, 0.5 and 0.7 and no break occurs in the coefficients. They show the local asymptotic power functions to be almost identical. Figure S.2 presents the local asymptotic power functions of the $\sup LR_{2,T}$ test when it accounts for a coefficient break at $\lambda^{c0} = 0.3$, 0.5 or 0.7. It also shows, the local asymptotic power functions of the $CUSQ$ test under the (correct) assumption of no break in the coefficients. Hence, this simulation design gives an advantage to the $CUSQ$ test and some power loss for the $\sup LR_{2,T}$ test might be expected. Indeed, the power of the $\sup LR_{2,T}$ test is slightly lower when the variance and the coefficient break dates coincide. This is because the permissible break dates around the true break date are not considered due to the concurrent nuisance break. However, the power loss of the $\sup LR_{2,T}$ test is very minor even though the $\sup LR_{2,T}$ test allows for a coefficient break. The power functions of both tests are almost identical when the two breaks are far apart. i.e., the case of $(\lambda^{v0}, \lambda^{c0}) = (0.3, 0.7)$ and $(0.7, 0.3)$.

Proof of Lemma S.1 The $\sup LR_{2,T}$ test is:

$$\begin{aligned} \sup LR_{2,T} &= 2[\log \hat{L}_T(\widetilde{T}^c; \widetilde{T}^v) - \log \widetilde{L}_T(\hat{T}^c)] \\ &= T \log \widetilde{\sigma}^2 - (T - \widetilde{T}^v) \log \hat{\sigma}_2^2 - \widetilde{T}^v \log \hat{\sigma}_1^2 \end{aligned}$$

where $\widetilde{\sigma}^2 = T^{-1} \sum_{t=1}^{T} (y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta}_{t,j})^2$, $\hat{\sigma}_2^2 = (T - \widetilde{T}^v)^{-1} \sum_{t=\widetilde{T}^v+1}^{T} (y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j})^2$ and $\hat{\sigma}_1^2 = \widetilde{T}^{v-1} \sum_{t=1}^{\widetilde{T}^v} (y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j})^2$. Applying a Taylor expansion to $\log \widetilde{\sigma}^2$, $\log \hat{\sigma}_2^2$ and $\log \hat{\sigma}_1^2$ around $\log \sigma_0^2$ (without loss of generality, let $\sigma_0^2 = \sigma_{20}^2$), we obtain

$$\sup LR_{2,T} = (F_{1,T} + F_{2,T}) + o_p(1)$$

where

$$\begin{aligned} F_{1,T} &= (\sigma_0^2)^{-1} [T\widetilde{\sigma}^2 - (T - \widetilde{T}^v)\hat{\sigma}_2^2 - \widetilde{T}^v \hat{\sigma}_1^2] \\ &= (\sigma_0^2)^{-1} \sum_{t=1}^{\widetilde{T}^v} \left[(y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta}_{t,j})^2 - (y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j})^2 \right] \\ &\quad + (\sigma_0^2)^{-1} \sum_{t=\widetilde{T}^v+1}^{T} \left[(y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta}_{t,j})^2 - (y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j})^2 \right] \end{aligned}$$

and

$$\begin{aligned} F_{2,T} &= -\frac{1}{2} \left[T \left(\frac{\widetilde{\sigma}^2 - \sigma_0^2}{\sigma_0^2} \right)^2 - (T - \widetilde{T}^v) \left(\frac{\hat{\sigma}_2^2 - \sigma_0^2}{\sigma_0^2} \right)^2 - \widetilde{T}^v \left(\frac{\hat{\sigma}_1^2 - \sigma_0^2}{\sigma_0^2} \right)^2 \right] \\ &= -\frac{1}{2}(I - II - III). \end{aligned}$$

We first show that $F_{1,T} = o_p(1)$. From Assumption A1, for any partition $\widetilde{T}^v$, we have $X_i'X_i = O_p(T)$, $Z_i'Z_i = O_p(T)$, $X_i'Z_i = O_p(T)$, $X_i'U_i = O_p(T^{1/2})$ and $Z_i'U_i = O_p(T^{1/2})$ for $i = 1$ and 2. In addition, under Assumptions A1 and A3 in which the change in the coefficient is assumed to shrink at rate υ_T, we obtain $\beta - \hat{\beta} = O_p(T^{-1/2})$, $\delta_j^0 - \hat{\delta}_j = O_p(T^{-1/2})$, $\hat{\beta} - \widetilde{\beta} = o_p(T^{-1/2})$ and $\hat{\delta}_j - \widetilde{\delta}_j = o_p(T^{-1/2})$ for $j = 1$ and 2. Hence, $F_{1,T} = o_p(1)$ is shown by directly following the proof of Theorem 1(b). If there is no break in the coefficient ($\delta_1^0 = \delta_2^0$), we also obtain $\beta - \hat{\beta} = O_p(T^{-1/2})$, $\delta_j^0 - \hat{\delta}_j = O_p(T^{-1/2})$, $\hat{\beta} - \widetilde{\beta} = o_p(T^{-1/2})$ and $\hat{\delta}_j - \widetilde{\delta}_j = o_p(T^{-1/2})$ for $j = 1$ and 2 no matter where $\widetilde{T}^c$ is. Hence, $F_{1,T} = o_p(1)$.

For $F_{2,T}$, we slightly change the notation and express the change in variance as $\sigma_{i0} - \sigma_{20} = (\sigma_i^{**}/\sigma_{20})/\sqrt{T}$ for $i = 1, 2$. We also denote σ_{20} by σ_0 without loss of generality so that $\sigma_2^{**} = 0$ by construction. Then,

$$\begin{aligned}
\sigma_{i0} - \sigma_0 &= \frac{\sigma_i^{**}/\sigma_0}{\sqrt{T}} \\
\sigma_{i0}^2 &= \sigma_0^2 + 2\frac{\sigma_i^{**}}{\sqrt{T}} + \frac{(\sigma_i^{**}/\sigma_0)^2}{T} = \sigma_0^2\left(1 + 2\frac{\sigma_i^{**}}{\sqrt{T}} + O(T^{-1})\right)
\end{aligned}$$

or

$$\frac{1}{\sigma_0^2} = \frac{1}{\sigma_{i0}^2}\left(1 + 2\frac{\sigma_i^{**}}{\sqrt{T}} + O(T^{-1})\right). \tag{S.6}$$

For each of the three terms, we have

$$\begin{aligned}
\sqrt{I} &= T^{-1/2}\sum_{t=1}^{T}\left[\frac{(y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta}_{t,j})^2}{\sigma_0^2} - 1\right] \\
&= T^{-1/2}\sum_{t=1}^{T}(\frac{u_t^2}{\sigma_0^2} - 1) + o_p(1) \\
&= T^{-1/2}\sum_{t=1}^{T}(\frac{u_t^2}{\sigma_{i0}^2} - 1) + T^{-1}\sum_{t=1}^{T}(\frac{u_t^2}{\sigma_{i0}^2})2\sigma_1^{**}I(t \leq T_1^{v0}) + o_p(1) \\
&\Rightarrow \sqrt{\psi}W(1) + 2\lambda^{v0}\sigma_1^{**},
\end{aligned}$$

$$\begin{aligned}
\sqrt{II} &= \left(\frac{T - \widetilde{T}^v}{T}\right)^{-1/2} T^{-1/2}\sum_{t=\widetilde{T}^v+1}^{T}(\frac{u_t^2}{\sigma_0^2} - 1) + o_p(1) \\
&= \left(\frac{T - \widetilde{T}^v}{T}\right)^{-1/2}\left\{T^{-1/2}\sum_{t=1}^{T}(\frac{u_t^2}{\sigma_0^2} - 1) - T^{-1/2}\sum_{t=1}^{\widetilde{T}^v}(\frac{u_t^2}{\sigma_0^2} - 1)\right\} + o_p(1) \\
&= \left(\frac{T - \widetilde{T}^v}{T}\right)^{-1/2}\left\{T^{-1/2}\sum_{t=1}^{T}(\frac{u_t^2}{\sigma_{i0}^2} - 1) - T^{-1/2}\sum_{t=1}^{\widetilde{T}^v}(\frac{u_t^2}{\sigma_{i0}^2} - 1)\right.
\end{aligned}$$

$$+T^{-1}\sum_{t=1}^{T}(\frac{u_t^2}{\sigma_{i0}^2})\sigma_1^{**}I(t\le T^{v0})-T^{-1}\sum_{t=1}^{\widetilde{T}^v}(\frac{u_t^2}{\sigma_{i0}^2})2\sigma_1^{**}I(t\le T^{v0})\Bigg\}+o_p(1)$$

$$\Rightarrow\quad \sqrt{\psi}\frac{W(1)-W(\lambda^v)}{\sqrt{1-\lambda^v}}+\frac{\lambda^{v0}-\min\{\lambda^{v0},\lambda^v\}}{\sqrt{1-\lambda^v}}2\sigma_1^{**},$$

$$\sqrt{III}=\left(\frac{\widetilde{T}^v}{T}\right)^{-1/2}\left\{T^{-1/2}\sum_{t=1}^{\widetilde{T}^v}(\frac{u_t^2}{\sigma_0^2}-1)\right\}+o_p(1)$$

$$=\quad\left(\frac{\widetilde{T}^v}{T}\right)^{-1/2}\left\{T^{-1/2}\sum_{t=1}^{\widetilde{T}^v}(\frac{u_t^2}{\sigma_{i0}^2}-1)+T^{-1}\sum_{t=1}^{\widetilde{T}^v}(\frac{u_t^2}{\sigma_{i0}^2})2\sigma_1^{**}I(t\le T_1^{v0})\right\}+o_p(1)$$

$$\Rightarrow\quad \sqrt{\psi}\frac{W(\lambda^v)}{\sqrt{\lambda^v}}+\frac{\min\{\lambda^{v0},\lambda^v\}}{\sqrt{\lambda^v}}2\sigma_1^{**}.$$

Therefore,

$$F_{2,T}\Rightarrow-\frac{\psi}{2}\left[\frac{\lambda^vW(1)-W(\lambda^v)}{\sqrt{\lambda^v(1-\lambda^v)}}+\frac{2}{\sqrt{\psi}}\sigma_1^{**}b(\lambda^v)\right]^2$$

where

$$b(\lambda^v)=\begin{cases}\lambda^{v0}\sqrt{\frac{1-\lambda^v}{\lambda^v}} & \text{if } \lambda^{v0}\le\lambda^v\\ (1-\lambda^{v0})\sqrt{\frac{\lambda^v}{1-\lambda^v}} & \text{if } \lambda^{v0}>\lambda^v\end{cases}.$$

This yields

$$\begin{aligned}\sup LR_{2,T}&\Rightarrow\sup_{\lambda^v\in\Lambda_{v,\varepsilon}^c}\frac{\psi}{2}[J(\lambda^v)]^2\\ J(\lambda^v)&=\frac{\lambda^vW(1)-W(\lambda^v)}{\sqrt{\lambda^v(1-\lambda^v)}}+\frac{2\sigma_1^{**}}{\sqrt{\psi}}b(\lambda^v)\\ b(\lambda^v)&=\begin{cases}\lambda^{v0}\sqrt{\frac{1-\lambda^v}{\lambda^v}} & \text{if } \lambda^{v0}\le\lambda^v\\ (1-\lambda^{v0})\sqrt{\frac{\lambda^v}{1-\lambda^v}} & \text{if } \lambda^{v0}>\lambda^v\end{cases}.\end{aligned}$$

The results for the $\sup LR_{1,T}$ follow because $F_{1,T}=o_p(1)$ holds also when there is no break in the coefficients.

Proof of Lemma S.2. The $\sup LR_{3,T}$ test is:

$$\begin{aligned}&\sup LR_{3,T}=2[\log\hat{L}_T(\widetilde{T}^c;\widetilde{T}^v)-\log\widetilde{L}_T(\hat{T}^v)]\\ =\;&(T-\hat{T}^v)\log\widetilde{\sigma}_2^2+\hat{T}^v\log\widetilde{\sigma}_1^2-(T-\widetilde{T}^v)\log\hat{\sigma}_2^2-\widetilde{T}^v\log\hat{\sigma}_1^2\end{aligned}$$

where $\widetilde{\sigma}_2^2 = (T - \widetilde{T}^v)^{-1}\sum_{t=\widetilde{T}^v+1}^{T}(y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta})^2$, $\widetilde{\sigma}_1^2 = (\widetilde{T}^v)^{-1}\sum_{t=1}^{\widetilde{T}^v}(y_t - x_t'\widetilde{\beta} - z_t'\widetilde{\delta})^2$, $\hat{\sigma}_2^2 = (T - \widetilde{T}^v)^{-1}\sum_{t=\widetilde{T}^v+1}^{T}(y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j})^2$ and $\hat{\sigma}_1^2 = (\widetilde{T}^v)^{-1}\sum_{t=1}^{\widetilde{T}^v}(y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_{t,j})^2$. Applying a Taylor expansion to $\log\widetilde{\sigma}_2^2$ and $\log\hat{\sigma}_2^2$ around $\log\sigma_{20}^2$, and to $\log\widetilde{\sigma}_1^2$ and $\log\hat{\sigma}_1^2$ around $\log\sigma_{10}^2$,

$$\sup LR_{3,T} = (F_{1,T} + F_{2,T}) + o_p(1) \tag{S.7}$$

where

$$F_{1,T} = (T - \hat{T}^v)\frac{\widetilde{\sigma}_2^2}{\sigma_{20}^2} - (T - \widetilde{T}^v)\frac{\hat{\sigma}_2^2}{\sigma_{20}^2} + \hat{T}^v\frac{\widetilde{\sigma}_1^2}{\sigma_{10}^2} - \widetilde{T}^v\frac{\hat{\sigma}_1^2}{\sigma_{10}^2}$$

and

$$\begin{aligned} F_{2,T} &= -\frac{1}{2}\left[(T - \hat{T}^v)\left(\frac{\widetilde{\sigma}_2^2 - \sigma_{20}^2}{\sigma_{20}^2}\right)^2 - (T - \widetilde{T}^v)\left(\frac{\hat{\sigma}_2^2 - \sigma_{20}^2}{\sigma_{20}^2}\right)^2\right] \\ &\quad -\frac{1}{2}\left[\hat{T}^v\left(\frac{\widetilde{\sigma}_1^2 - \sigma_{10}^2}{\sigma_{10}^2}\right)^2 - \widetilde{T}^v\left(\frac{\hat{\sigma}_1^2 - \sigma_{10}^2}{\sigma_{10}^2}\right)^2\right]. \end{aligned}$$

We first show that $F_{2,T} = o_p(1)$. We have

$$\begin{aligned} F_{2,T} &= -\frac{1}{2}\left[\frac{T - \hat{T}^v}{T}\left[\sqrt{T}\left(\frac{\widetilde{\sigma}_2^2 - \sigma_{20}^2}{\sigma_{20}^2}\right)\right]^2 - \frac{T - \widetilde{T}^v}{T}\left[\sqrt{T}\left(\frac{\hat{\sigma}_2^2 - \sigma_{20}^2}{\sigma_{20}^2}\right)\right]^2\right] \\ &\quad -\frac{1}{2}\left[\frac{\hat{T}^v}{T}\left[\sqrt{T}\left(\frac{\widetilde{\sigma}_1^2 - \sigma_{10}^2}{\sigma_{10}^2}\right)\right]^2 - \frac{\widetilde{T}^v}{T}\left[\sqrt{T}\left(\frac{\hat{\sigma}_1^2 - \sigma_{10}^2}{\sigma_{10}^2}\right)\right]^2\right] \end{aligned}$$

where $[(T - \hat{T}^v)/T][\sqrt{T}(\widetilde{\sigma}_2^2 - \sigma_{20}^2)/\sigma_{20}^2]^2$ and $[(T - \widetilde{T}^v)/T][\sqrt{T}(\hat{\sigma}_2^2 - \sigma_{20}^2)/\sigma_{20}^2]^2$ have the same limit distribution and $(\hat{T}^v/T)[\sqrt{T}(\widetilde{\sigma}_1^2 - \sigma_{10}^2)/\sigma_{10}^2]^2$ and $(\widetilde{T}^v/T)[\sqrt{T}(\hat{\sigma}_2^2 - \sigma_{20}^2)/\sigma_{20}^2]^2$ have the same limit distribution under Assumption A3. These also hold when there is no break in the variance. For $F_{1,T}$, let $\sigma_0 = \sigma_{20}$ without loss of generality, then

$$\begin{aligned} F_{1,T} &= (\sigma_0^2)^{-1}\left[(T - \hat{T}^v)\widetilde{\sigma}_2^2 - (T - \widetilde{T}^v)\hat{\sigma}_2^2 + \hat{T}^v\widetilde{\sigma}_1^2 - \widetilde{T}^v\hat{\sigma}_1^2\right] \\ &\quad -(\sigma_0^2)^{-1}\left(\frac{\sigma_{10}^2 - \sigma_0^2}{\sigma_{10}^2}\right)(\hat{T}^v\widetilde{\sigma}_1^2 - \widetilde{T}^v\hat{\sigma}_1^2). \end{aligned}$$

The first term becomes,

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

and the second term is $o_p(1)$ under Assumption A3. When there is no break in variance, the second term is zero because $\sigma_{10}^2 - \sigma_0^2 = \sigma_{10}^2 - \sigma_{20}^2 = 0$. Then,

$$
\begin{aligned}
& D^r - D^u(2) - D^u(1) \\
= & -\left[T^{-1/2}\sum_{t=1}^{T}(u_t z_t' + I(t \leq T^{c0})\delta^{*\prime} z_t z_t')\right]\left(T^{-1}\sum_{t=1}^{T} z_t z_t'\right) \\
& \times \left[T^{-1/2}\sum_{t=1}^{T}(z_t u_t + z_t z_t'\delta^* I(t \leq T^{c0}))\right] \\
& + \left[T^{-1/2}\sum_{t=1}^{\widetilde{T}^c}(u_t z_t' + I(t \leq T^{c0})\delta^{*\prime} z_t z_t')\right]\left(T^{-1}\sum_{t=1}^{\widetilde{T}^c} z_t z_t'\right) \\
& \times \left[T^{-1/2}\sum_{t=1}^{\widetilde{T}^c}(z_t u_t + z_t z_t'\delta^* I(t \leq T^{c0}))\right] \\
& + \left[T^{-1/2}\sum_{t=\widetilde{T}^c+1}^{T}(u_t z_t' + I(t \leq T^{c0})\delta^{*\prime} z_t z_t')\right]\left(T^{-1}\sum_{t=\widetilde{T}^c+1}^{T} z_t z_t'\right) \\
& \times \left[T^{-1/2}\sum_{t=\widetilde{T}^c+1}^{T}(z_t u_t + z_t z_t'\delta^* I(t \leq T^{c0}))\right] + o_p(1) \\
\Rightarrow & J(\lambda^c)' J(\lambda^c)
\end{aligned}
$$

where

$$
\begin{aligned}
J(\lambda^c) &= \left(\frac{\lambda^c W_q(1) - W_q(\lambda^c)}{\sqrt{\lambda^c(1-\lambda^c)}}\right) + Q^{1/2}\delta^* b(\lambda^c) \\
b(\lambda^c) &= \begin{cases} \lambda^{c0}\sqrt{\frac{1-\lambda^c}{\lambda^c}} & \text{if } \lambda^{c0} \leq \lambda^c \\ (1-\lambda^{c0})\sqrt{\frac{\lambda^c}{1-\lambda^c}} & \text{if } \lambda^{c0} > \lambda^c \end{cases}
\end{aligned}
$$

and $Q \equiv p\lim_{T\to\infty}(T^{-1}\sum_{t=1}^{T} z_t z_t')$. Hence, from (S.7) the results of the $\sup LR_{3,T}$ test is obtained. The result for the the $\sup LR_T$ is also obtained since we showed $F_{2,T} = o_p(1)$ to hold when there is no variance break.

Proof of Lemma S.3. By (S.6), we obtain

$$\sigma_{i0}^2/\sigma_0^2 = 1 + 2\sigma_i^{**}/\sqrt{T} + O(T^{-1})$$

and the test statistic $CUSQ$ is such that:

$$
\begin{aligned}
CUSQ &= \sup_{\lambda^v \in [0,1]}\left|T^{-1/2}\left[\sum_{t=1}^{[T\lambda^v]}\left(\frac{u_t^2}{\sigma_{i0}^2}\frac{\sigma_{i0}^2}{\sigma_0^2} - 1\right) - \frac{[T\lambda^v]}{T}\sum_{t=1}^{T}\left(\frac{u_t^2}{\sigma_{i0}^2}\frac{\sigma_{i0}^2}{\sigma_0^2} - 1\right)\right]\right| + o_p(1) \\
&= \sup_{\lambda^v \in [0,1]}\left|T^{-1/2}\left[\sum_{t=1}^{[T\lambda^v]}\left\{\frac{u_t^2}{\sigma_{i0}^2}\left(1 + 2\frac{\sigma_1^{**}}{\sqrt{T}}I(t \leq T^{v0})\right) - 1\right\}\right.\right. \\
&\quad \left.\left. - \frac{[T\lambda^v]}{T}\sum_{t=1}^{T}\left\{\frac{u_t^2}{\sigma_{i0}^2}\left(1 + 2\frac{\sigma_1^{**}}{\sqrt{T}}I(t \leq T^{v0})\right) - 1\right\}\right]\right| + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
&= \sup_{\lambda^v \in [0,1]} \Bigg| T^{-1/2} \left[\sum_{t=1}^{[T\lambda^v]} \left(\frac{u_t^2}{\sigma_{i0}^2} - 1 \right) - \frac{[T\lambda^v]}{T} \sum_{t=1}^{T} \left(\frac{u_t^2}{\sigma_{i0}^2} - 1 \right) \right] \\
&\quad + T^{-1} \left[\sum_{t=1}^{[T\lambda^v]} \left(\frac{u_t^2}{\sigma_{i0}^2} \right) 2\sigma_1^{**} I(t \le T^{v0}) - \frac{[T\lambda^v]}{T} \sum_{t=1}^{T} \left(\frac{u_t^2}{\sigma_{i0}^2} \right) 2\sigma_1^{**} I(t \le T^{v0}) \right] \Bigg| + o_p(1) \\
&\Rightarrow \sqrt{\psi} \sup_{\lambda^v \in [0,1]} \left| [W(\lambda^v) - \lambda^v W(1)] + \frac{2\sigma_1^{**}}{\sqrt{\psi}} \mu(\lambda^v) \right|
\end{aligned}
$$

where

$$
\mu(\lambda^v) = \begin{cases} \lambda^{v0}(1-\lambda^v) & \text{if } \lambda^{v0} \le \lambda^v \\ \lambda^v(1-\lambda^{v0}) & \text{if } \lambda^{v0} > \lambda^v \end{cases} .
$$

E: Robustness to non-normal errors.

Given that our tests are based on a quasi-likelihood framework assuming normal errors, it is useful to assess their size and power under non-normal error distributions. We focus on the tests for the structural changes in variance, i.e. the $\sup LR_{1,T}$ and $\sup LR_{2,T}$ tests, mostly because these are the most prone to be affected by non-normality; e.g., the test for a single coefficient break coincides with that derived by Andrews (1993) using a GMM-based approach. To investigate the $\sup LR_{1,T}$ test, we generate the same data as the experiment pertaining to Table 2:

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

where $h_0 = \tau_1 / (1 - \gamma - \rho)$ and u_t is drawn from the following well-known non-normal distributions: (a) the t distribution with 5 degrees of freedom (t_5), (b) a mixture of two normal distributions: $v_1 I(z \le 0.5) + v_2 I(z > 0.5)$, where $z \sim U[0,1]$, $v_1 \sim N(-1,1)$ and $v_2 \sim N(1,1)$ (c) the χ^2 distribution with 5 degrees of freedom and (d) an exponential distribution $-\ln(v)$, $v \sim U[0,1]$. These distributions were chosen as empirically relevant examples following Bai and Ng (2005). To facilitate comparisons, the errors are normalized by subtracting the sample mean and dividing by the sample standard deviation of each Monte Carlo repetition. The model parameter values are set at $c = 0.5, \tau_1 = 0.1, \rho = 0.2$, and $\varepsilon = 0.15$. We consider $\alpha = 0.2, 0.7$ and $\gamma = 0.1, 0.3, 0.5$. The sample size is $T = 100, 200$. Table S.4 presents the exact size and power of the $\sup LR_{1,T}$, $UD\max_{1,T}$ and $CUSQ$ tests. The reported values are roughly comparable to those with normal errors in Table 2, i.e., little if any size distortions. In all cases, the power decreases to some extent. Note, however, that this is also the case for the $CUSQ$ test and the relative advantage of the $\sup LR_{1,T}$ and $UD\max_{1,T}$ tests over the $CUSQ$ test remains under these non-normal errors.

For the $\sup LR_{2,T}$ test, we use the same data generating process as that corresponding to Table 4 to assess the size, i.e.,

$$y_t = \mu_1 + \mu_2 1(t > [0.5T]) + e_t, \tag{S.8}$$

with e_t drawn from one of the four types of non-normal distributions. Again, the errors are standardized to have mean zero and variance one in each Monte Carlo repetition. We set $\mu_1 = 0$ and the truncation $\varepsilon = 0.15$, although we obtained similar results for other choices of ε. Table S.5 shows that the size distortions are minor in all cases. To assess the power of the $\sup LR_{2,T}$ test, we use again DGP (S.8). The errors are standardized to have mean zero and variance one when $t \leq [0.25T]$ and $1+\theta$ when $t > [0.25T]$ so that θ indicates the magnitude of the break. The results are presented in Table S.6. Relative to the results in Table 5, we have, as expected, some power reductions. The extent of the power losses vary across the different distributions. Nevertheless, the test remains informative.

F: Size and Power of the $\sup LR^*_{1,T}$ test in the case of normal errors.

Table S.7 presents results related to Table 2 for the statistic $\sup LR^*_{1,T}$ when testing for a single break in variance assuming no break in regression coefficients but with the static model and normal errors. The DGP is $y_t = e_t$ with $e_t \sim i.i.d.N(0, 1+\theta I(t > [.5T])$ and θ varies between 0 and 1.5. The trimming parameter is set to $\varepsilon = 0.15$. The results show that using $\hat{\psi}$ with the full correction yields power and exact size similar to tests with a correction that correctly assumes $i.i.d.$ errors, though here imposing normality can lead to tests with somewhat higher power. This confirms that using the full correction entails little power loss or size distortions.

We also investigate the findings that the $UD\max LR_{1,T}$ test can have power close to that of $\sup LR^*_{1,T}$ under a single break model even though the former considers a wider range of alternatives by using a simple design with normal errors. We also compare them with the $CUSQ$ test described in the main text. We use the same DGP and the results are presented in Table S.8. They show the exact sizes of the $\sup LR^*_{1,T}$ and $UD\max LR_{1,T}$ tests to be close to the nominal 5% size. The $CUSQ$ test is slightly undersized. The power functions of the three tests are very close.

G: Size of the $\sup LR^*_{4,T}$ and $UD\max$ tests in the case of normal errors.

We present results about the size properties of the $\sup LR^*_{4,T}$ and $UD\max$ tests with normal *i.i.d.* errors, with the DGP set to $y_t = e_t \sim i.i.d.\ N(0,1)$. We use three values of the trimming parameter $\varepsilon = 0.1, 0.15$ and 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) $m_a = 2$, $n_a = 1$. Two sample sizes are used, $T = 100$, 200. The results are presented in Table S.9 and they show the size to be slightly conservative, as expected since the critical values are from a limit distribution that provides an upper bound. Nevertheless, the size is close to the nominal 5% level in every case.

H: Split-sample method to select the number of breaks.

We present results for a split-sample method to estimate the number of breaks in δ and σ^2. It is based on a specific to general sequential procedure which is a modification of the sequential procedure discussed in Qu and Perron (2007). Our problem is, however, more complex since we wish to ascertain what types of break occur at any given selected break date, not only to know whether some kind of break did occur. Hence, the need for some refinements. The starting point is to consider the testing problem for the number of breaks in the union of the coefficients and variance breaks K. This is implemented by using the $Seq_T(l+1|l)$ test proposed by Qu and Perron (2007). The next step is to decide whether a break in coefficients, in variance or in both has occurred at each of the selected break dates. We then perform standard hypothesis testing for the equality of the parameters across adjacent segments. Since the limit distribution of the estimates of the parameters of the model are the same whether using estimates of the break dates or their true value, standard procedures can be applied. Consider first the case of testing whether the regression coefficients are equal across the two regimes $(\hat{T}_{k-1}, \hat{T}_k)$, regime k, and $(\hat{T}_k, \hat{T}_{k+1})$, regime $k+1$, separated by the k^{th} break $(k = 1, ..., K)$. Denote the true value of the regression coefficients in regimes k and $k+1$ by δ_k and δ_{k+1}, respectively. The null and alternative hypotheses are $H_0 : \delta_k = \delta_{k+1}$ and $H_1 : \delta_k \neq \delta_{k+1}$. Note that since there is a break in δ and/or σ^2, under H_0 there must be a change in σ^2. Hence, the test to be applied is a standard Chow-type test allowing for a change in variance across regimes (see Goldfeld and Quandt, 1978). Consider now the testing problem $H_0 : \sigma^2_k = \sigma^2_{k+1}$ versus $H_1 : \sigma^2_k \neq \sigma^2_{k+1}$, where σ^2_k and σ^2_{k+1} are the error variances in regimes k and $k+1$, respectively. The Wald test is

$$W_k = \frac{(\hat{T}_k - \hat{T}_{k-1})(\hat{T}_{k+1} - \hat{T}_k)}{(\hat{T}_{k+1} - \hat{T}_{k-1})(\hat{\mu}_4 - \hat{\sigma}^4)} \left(\hat{\sigma}^2_{k+1} - \hat{\sigma}^2_k\right)^2,$$

where $\hat{\sigma}^2_k$ and $\hat{\sigma}^2_{k+1}$ are the MLE of σ^2_k and σ^2_{k+1} (constructed allowing δ to be different in regimes k and $k+1$), and $\hat{\mu}_4$ is a consistent estimate of $E(u_t^4)$, e.g., $\hat{\mu}_4 = (\hat{T}_{k+1} - \hat{T}_{k-1})^{-1} \sum_{\hat{T}_{k-1}+1}^{\hat{T}_{k+1}} \hat{u}_t^4$, constructed under H_1 to maximize power. The simulation design is the same as stated in Section 6. The results for this split-sample approach are presented in Table S.10. Of note, there are cases for which the probability of making the correct selection is quite low; e.g., when both changes in mean and variance are not large and occur at different dates, especially when they are far apart. The basic reason for that is that the sequential test of Qu and Perron (2007) jointly tests whether a break in both regression coefficients and variance occur. Hence, if only one type of break occurs the power can be quite low unless the magnitudes of the breaks are large. Unfortunately, this situation is expected to be quite common in practice (see, Perron and Yamamoto, 2019). Hence, though this procedure is valid in large samples, it should not be applied mechanically. Care must be exercised to assess whether we are in a situation where its finite sample properties are rather poor.

References

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

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

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

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

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

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

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

Table S.1: 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 $u_t \sim i.i.d.\ N(0,1)$, $h_t = \tau_1 + \gamma e^2_{t-1} + \rho h_{t-1}$, $h_0 = \tau_1/(1-\gamma-\rho)$, $\tau_1 = 1$, $T = 100$, $\varepsilon = 0.20$, 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 S.2: Power of the sup $LR^*_{4,T}$ using different estimates of ψ in the case of GARCH(1) errors (DGP: $y_t = \mu_1 + \mu_2 1(t > [0.25T]) + e_t$, $e_t = u_t\sqrt{h_t}$, with $u_t \sim i.i.d.\ N(0,1)$, $h_t = \tau_1 + \tau_2 1(t > [0.75T]) + \gamma e^2_{t-1} + \rho h_{t-1}$, $h_0 = \tau_1/(1-\gamma-\rho)$, $\tau_1 = 1$, $\rho = 0.2$, $T = 100$; $\varepsilon = 0.20$).

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 S.3: 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 (8); 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 S.4: Size and power of the $supLR_{1,T}(n_a = 1, \varepsilon)$, $UD\max LR_{1,T}$ and $CUSQ$ tests in a dynamic model with GARCH(1,1) errors

(a) t_5 distribution

	$T = 100$																	
	$\alpha = 0.2$									$\alpha = 0.7$								
	$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$			$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$		
τ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.041	0.039	0.026	0.064	0.064	0.025	0.094	0.090	0.029	0.046	0.038	0.023	0.065	0.062	0.028	0.091	0.085	0.031
0.05	0.094	0.090	0.068	0.101	0.098	0.053	0.108	0.106	0.047	0.089	0.079	0.064	0.088	0.093	0.047	0.104	0.103	0.037
0.1	0.203	0.190	0.182	0.184	0.187	0.137	0.178	0.179	0.102	0.194	0.189	0.180	0.188	0.187	0.130	0.169	0.175	0.096
0.15	0.327	0.303	0.273	0.273	0.262	0.193	0.232	0.218	0.134	0.310	0.290	0.277	0.264	0.251	0.198	0.233	0.226	0.134
0.2	0.435	0.415	0.390	0.372	0.364	0.268	0.300	0.296	0.183	0.431	0.410	0.385	0.366	0.357	0.258	0.306	0.291	0.181
0.3	0.620	0.600	0.530	0.501	0.490	0.385	0.407	0.399	0.259	0.615	0.594	0.524	0.494	0.479	0.386	0.394	0.380	0.259
	$T = 200$																	
	$\alpha = 0.2$									$\alpha = 0.7$								
	$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$			$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$		
τ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.041	0.039	0.027	0.054	0.054	0.036	0.057	0.052	0.040	0.042	0.035	0.029	0.051	0.052	0.038	0.053	0.048	0.039
0.05	0.129	0.123	0.136	0.113	0.104	0.101	0.089	0.085	0.073	0.127	0.121	0.142	0.108	0.102	0.103	0.086	0.081	0.068
0.1	0.358	0.340	0.369	0.279	0.262	0.258	0.229	0.207	0.169	0.367	0.347	0.375	0.275	0.248	0.254	0.213	0.194	0.169
0.15	0.560	0.536	0.565	0.420	0.403	0.388	0.294	0.279	0.257	0.557	0.535	0.563	0.403	0.387	0.383	0.290	0.269	0.250
0.2	0.718	0.705	0.712	0.551	0.533	0.523	0.397	0.380	0.320	0.707	0.690	0.708	0.543	0.521	0.519	0.387	0.379	0.309
0.3	0.845	0.837	0.825	0.700	0.685	0.636	0.519	0.507	0.435	0.844	0.832	0.827	0.696	0.682	0.636	0.512	0.490	0.431

(b) mixture of normal distributions

	$T=100$																	
	$\alpha=0.2$									$\alpha=0.7$								
	$\gamma=0.1$			$\gamma=0.3$			$\gamma=0.5$			$\gamma=0.1$			$\gamma=0.3$			$\gamma=0.5$		
τ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.102	0.105	0.037	0.107	0.117	0.041	0.110	0.121	0.042	0.067	0.062	0.026	0.109	0.109	0.036	0.112	0.119	0.049
0.05	0.184	0.186	0.128	0.141	0.156	0.085	0.143	0.157	0.086	0.213	0.207	0.188	0.191	0.184	0.135	0.145	0.151	0.082
0.1	0.350	0.343	0.268	0.241	0.255	0.146	0.231	0.246	0.135	0.477	0.462	0.452	0.349	0.335	0.265	0.224	0.236	0.142
0.15	0.522	0.509	0.429	0.335	0.325	0.216	0.329	0.323	0.218	0.714	0.695	0.703	0.510	0.493	0.418	0.320	0.309	0.211
0.2	0.639	0.632	0.552	0.428	0.430	0.285	0.439	0.446	0.287	0.851	0.840	0.828	0.637	0.628	0.533	0.423	0.413	0.273
0.3	0.785	0.769	0.677	0.581	0.573	0.389	0.555	0.544	0.386	0.949	0.939	0.906	0.783	0.763	0.664	0.558	0.546	0.393
	$T=200$																	
	$\alpha=0.2$									$\alpha=0.7$								
	$\gamma=0.1$			$\gamma=0.3$			$\gamma=0.5$			$\gamma=0.1$			$\gamma=0.3$			$\gamma=0.5$		
τ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.059	0.062	0.037	0.065	0.068	0.042	0.051	0.050	0.043	0.059	0.058	0.029	0.063	0.067	0.040	0.051	0.049	0.041
0.05	0.373	0.359	0.411	0.243	0.236	0.241	0.143	0.137	0.126	0.364	0.348	0.408	0.241	0.231	0.236	0.141	0.132	0.116
0.1	0.831	0.822	0.848	0.556	0.543	0.566	0.307	0.292	0.272	0.827	0.815	0.848	0.548	0.538	0.558	0.296	0.286	0.265
0.15	0.968	0.962	0.973	0.758	0.750	0.761	0.455	0.441	0.399	0.965	0.960	0.974	0.754	0.744	0.747	0.438	0.426	0.389
0.2	0.995	0.995	0.996	0.867	0.864	0.864	0.584	0.575	0.528	0.996	0.996	0.995	0.869	0.863	0.868	0.577	0.567	0.527
0.3	1.000	1.000	0.999	0.954	0.948	0.930	0.726	0.702	0.634	1.000	1.000	0.999	0.951	0.944	0.923	0.703	0.690	0.629

(c) χ^2_5 distribution

	$T=100$																	
	$\alpha=0.2$									$\alpha=0.7$								
	$\gamma=0.1$			$\gamma=0.3$			$\gamma=0.5$			$\gamma=0.1$			$\gamma=0.3$			$\gamma=0.5$		
τ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.024	0.025	0.015	0.054	0.050	0.029	0.087	0.092	0.041	0.025	0.025	0.015	0.051	0.048	0.026	0.081	0.087	0.044
0.05	0.085	0.080	0.068	0.095	0.092	0.059	0.111	0.116	0.055	0.083	0.073	0.069	0.091	0.088	0.058	0.101	0.112	0.050
0.1	0.191	0.178	0.169	0.188	0.177	0.130	0.170	0.176	0.099	0.195	0.187	0.159	0.180	0.182	0.135	0.169	0.170	0.092
0.15	0.359	0.329	0.309	0.307	0.293	0.227	0.260	0.254	0.153	0.356	0.329	0.304	0.302	0.282	0.223	0.251	0.259	0.153
0.2	0.510	0.490	0.448	0.427	0.414	0.333	0.334	0.324	0.213	0.509	0.487	0.443	0.408	0.391	0.315	0.320	0.321	0.208
0.3	0.637	0.614	0.573	0.548	0.536	0.416	0.434	0.432	0.284	0.628	0.616	0.559	0.543	0.522	0.407	0.426	0.409	0.280
	$T=200$																	
	$\alpha=0.2$									$\alpha=0.7$								
	$\gamma=0.1$			$\gamma=0.3$			$\gamma=0.5$			$\gamma=0.1$			$\gamma=0.3$			$\gamma=0.5$		
τ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.031	0.028	0.025	0.044	0.041	0.031	0.053	0.053	0.042	0.032	0.032	0.029	0.044	0.043	0.035	0.050	0.050	0.049
0.05	0.140	0.134	0.133	0.136	0.124	0.116	0.118	0.109	0.085	0.138	0.129	0.135	0.130	0.124	0.107	0.109	0.102	0.084
0.1	0.388	0.366	0.415	0.282	0.259	0.260	0.196	0.182	0.145	0.370	0.348	0.384	0.270	0.247	0.261	0.187	0.178	0.157
0.15	0.623	0.603	0.633	0.478	0.455	0.438	0.333	0.326	0.260	0.645	0.623	0.650	0.466	0.447	0.429	0.325	0.310	0.256
0.2	0.781	0.761	0.790	0.601	0.593	0.576	0.415	0.404	0.329	0.765	0.744	0.766	0.592	0.581	0.560	0.408	0.393	0.326
0.3	0.926	0.915	0.911	0.786	0.775	0.732	0.589	0.573	0.493	0.911	0.901	0.895	0.785	0.766	0.734	0.575	0.564	0.470

(d) exponential distribution

	$T = 100$																	
	$\alpha = 0.2$									$\alpha = 0.7$								
	$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$			$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$		
τ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.032	0.029	0.018	0.047	0.044	0.028	0.062	0.066	0.032	0.034	0.029	0.018	0.042	0.039	0.024	0.057	0.058	0.028
0.05	0.067	0.061	0.051	0.093	0.088	0.056	0.110	0.110	0.060	0.069	0.062	0.046	0.097	0.092	0.063	0.114	0.116	0.058
0.1	0.129	0.117	0.075	0.140	0.132	0.081	0.148	0.143	0.073	0.122	0.118	0.070	0.121	0.120	0.077	0.137	0.134	0.070
0.15	0.218	0.200	0.153	0.215	0.203	0.138	0.227	0.212	0.118	0.220	0.202	0.157	0.211	0.199	0.137	0.209	0.198	0.110
0.2	0.274	0.257	0.198	0.259	0.255	0.161	0.254	0.250	0.139	0.275	0.255	0.209	0.262	0.250	0.165	0.249	0.240	0.128
0.3	0.445	0.419	0.334	0.407	0.393	0.259	0.374	0.363	0.231	0.440	0.409	0.328	0.386	0.372	0.249	0.355	0.340	0.215
	$T = 200$																	
	$\alpha = 0.2$									$\alpha = 0.7$								
	$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$			$\gamma = 0.1$			$\gamma = 0.3$			$\gamma = 0.5$		
τ_2	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ	LR	UDmax	CUSQ
0	0.029	0.025	0.025	0.044	0.041	0.036	0.058	0.055	0.036	0.031	0.024	0.026	0.040	0.039	0.034	0.050	0.046	0.034
0.05	0.099	0.090	0.089	0.098	0.089	0.083	0.114	0.104	0.083	0.091	0.082	0.090	0.094	0.084	0.080	0.105	0.095	0.081
0.1	0.232	0.214	0.226	0.202	0.188	0.178	0.188	0.177	0.135	0.232	0.215	0.228	0.202	0.184	0.178	0.185	0.173	0.131
0.15	0.364	0.350	0.355	0.307	0.303	0.262	0.267	0.252	0.201	0.370	0.348	0.347	0.307	0.291	0.261	0.254	0.240	0.194
0.2	0.539	0.517	0.512	0.452	0.437	0.409	0.380	0.360	0.291	0.537	0.508	0.515	0.444	0.428	0.395	0.358	0.341	0.281
0.3	0.717	0.701	0.676	0.593	0.584	0.525	0.478	0.469	0.365	0.717	0.699	0.667	0.583	0.579	0.523	0.465	0.453	0.366

Table S.5: Size of the sup $LR^*_{2,T}(m_a = 1, n_a = 1, \varepsilon | n = 0, m_a = 1)$ test under non-normal errors ($\varepsilon = 0.15$)

	$T = 100$				$T = 200$			
μ_2	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
0	0.020	0.039	0.033	0.047	0.021	0.031	0.039	0.042
0.1	0.021	0.035	0.036	0.031	0.024	0.030	0.032	0.035
0.25	0.014	0.040	0.029	0.039	0.021	0.034	0.034	0.027
0.5	0.024	0.040	0.032	0.033	0.015	0.032	0.031	0.032
0.75	0.031	0.039	0.027	0.033	0.013	0.029	0.027	0.024
1	0.029	0.027	0.026	0.033	0.015	0.039	0.030	0.026
2	0.016	0.027	0.014	0.030	0.022	0.033	0.030	0.018
4	0.025	0.027	0.018	0.032	0.020	0.028	0.021	0.031
10	0.022	0.033	0.020	0.024	0.022	0.040	0.016	0.023
20	0.020	0.039	0.017	0.026	0.029	0.025	0.023	0.029

Note: (a) the t_5 distribution, (b) the mixture of normal distributions, (c) the χ^2_5 distribution., (d) the exponential distribution.

Table S.6: Power of the sup $LR^{*}_{2,T}(m_a = 1, n_a = 1, \varepsilon | n = 0, m_a = 1)$ test under non-normal errors ($\varepsilon = 0.1$)

	$T = 100$													
	(a) t_5 distribution							(b) mixture of normal distributions						
$\theta \backslash \mu_2$	0	0.1	0.5	2	4	10	20	0	0.1	0.5	2	4	10	20
0.25	0.028	0.034	0.027	0.034	0.033	0.029	0.025	0.073	0.063	0.049	0.060	0.059	0.089	0.080
0.5	0.039	0.043	0.040	0.055	0.037	0.054	0.047	0.117	0.121	0.111	0.137	0.117	0.124	0.155
0.75	0.063	0.065	0.077	0.064	0.079	0.072	0.069	0.196	0.202	0.200	0.210	0.213	0.249	0.258
1	0.075	0.107	0.091	0.120	0.112	0.106	0.107	0.334	0.295	0.278	0.304	0.326	0.320	0.323
1.25	0.132	0.135	0.147	0.161	0.159	0.125	0.135	0.418	0.393	0.402	0.434	0.460	0.474	0.450
1.5	0.186	0.168	0.193	0.159	0.200	0.199	0.188	0.510	0.489	0.473	0.479	0.553	0.541	0.586
2	0.259	0.245	0.252	0.305	0.297	0.297	0.281	0.668	0.677	0.667	0.701	0.722	0.750	0.771
3	0.395	0.407	0.389	0.431	0.437	0.430	0.447	0.862	0.890	0.881	0.885	0.921	0.938	0.913
4	0.535	0.543	0.567	0.559	0.560	0.599	0.593	0.949	0.966	0.954	0.963	0.974	0.976	0.975
	(c) χ^2_5 distribution							(d) exponential distribution						
$\theta \backslash \mu_2$	0	0.1	0.5	2	4	10	20	0	0.1	0.5	2	4	10	20
0.25	0.051	0.052	0.036	0.029	0.034	0.027	0.033	0.041	0.048	0.048	0.025	0.021	0.035	0.032
0.5	0.066	0.061	0.062	0.038	0.046	0.055	0.059	0.062	0.066	0.059	0.031	0.042	0.045	0.051
0.75	0.092	0.102	0.089	0.095	0.089	0.090	0.071	0.087	0.085	0.078	0.063	0.059	0.066	0.048
1	0.128	0.133	0.131	0.122	0.112	0.122	0.133	0.101	0.091	0.085	0.086	0.068	0.085	0.071
1.25	0.150	0.167	0.183	0.182	0.170	0.157	0.163	0.127	0.116	0.112	0.125	0.098	0.093	0.095
1.5	0.200	0.199	0.209	0.211	0.234	0.229	0.231	0.142	0.141	0.125	0.142	0.119	0.133	0.128
2	0.311	0.289	0.294	0.281	0.313	0.280	0.305	0.181	0.174	0.205	0.169	0.176	0.151	0.162
3	0.482	0.459	0.447	0.435	0.472	0.497	0.495	0.276	0.260	0.262	0.272	0.269	0.251	0.291
4	0.599	0.572	0.590	0.555	0.617	0.625	0.603	0.357	0.376	0.360	0.350	0.355	0.381	0.377

	$T = 200$													
	(a) t_5 distribution							(b) mixture of normal distributions						
$\theta \backslash \mu_2$	0	0.1	0.5	2	4	10	20	0	0.1	0.5	2	4	10	20
0.25	0.030	0.037	0.031	0.041	0.033	0.044	0.046	0.089	0.083	0.071	0.097	0.093	0.112	0.088
0.5	0.066	0.073	0.079	0.068	0.066	0.073	0.077	0.235	0.241	0.216	0.254	0.274	0.259	0.261
0.75	0.125	0.119	0.115	0.137	0.145	0.141	0.138	0.441	0.453	0.424	0.458	0.509	0.489	0.518
1	0.177	0.181	0.217	0.206	0.214	0.224	0.241	0.624	0.629	0.607	0.694	0.717	0.734	0.715
1.25	0.252	0.273	0.251	0.326	0.287	0.315	0.337	0.781	0.796	0.759	0.834	0.844	0.864	0.846
1.5	0.361	0.323	0.353	0.408	0.395	0.436	0.383	0.893	0.904	0.887	0.917	0.938	0.940	0.943
2	0.511	0.513	0.507	0.532	0.563	0.542	0.543	0.976	0.978	0.975	0.989	0.988	0.990	0.993
3	0.730	0.726	0.727	0.748	0.758	0.775	0.773	1.000	0.999	1.000	1.000	0.999	1.000	1.000
4	0.842	0.833	0.838	0.853	0.874	0.875	0.840	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(c) χ^2_5 distribution							(d) exponential distribution						
$\theta \backslash \mu_2$	0	0.1	0.5	2	4	10	20	0	0.1	0.5	2	4	10	20
0.25	0.052	0.041	0.051	0.034	0.041	0.044	0.040	0.045	0.054	0.035	0.035	0.032	0.023	0.028
0.5	0.097	0.114	0.090	0.083	0.091	0.091	0.084	0.074	0.058	0.077	0.052	0.052	0.061	0.042
0.75	0.174	0.164	0.182	0.133	0.152	0.159	0.197	0.105	0.105	0.099	0.076	0.075	0.076	0.104
1	0.258	0.239	0.250	0.248	0.235	0.252	0.253	0.147	0.154	0.131	0.135	0.135	0.115	0.147
1.25	0.341	0.354	0.322	0.337	0.339	0.363	0.361	0.196	0.185	0.194	0.178	0.178	0.184	0.186
1.5	0.461	0.432	0.410	0.409	0.442	0.450	0.452	0.226	0.232	0.231	0.258	0.258	0.237	0.226
2	0.597	0.599	0.598	0.627	0.598	0.601	0.608	0.314	0.342	0.337	0.332	0.332	0.362	0.337
3	0.803	0.767	0.789	0.817	0.822	0.822	0.816	0.512	0.517	0.539	0.532	0.532	0.541	0.503
4	0.900	0.903	0.889	0.917	0.916	0.915	0.912	0.644	0.698	0.652	0.668	0.668	0.664	0.692

Table S.7: Size and power of the sup $LR^*_{1,T}$ test using different corrections in the case of i.i.d. normal errors (DGP: $y_t = e_t$; $e_t \sim i.i.d.\ N(0, 1 + \theta 1(t > [.5T]))$, $\varepsilon = 0.15$)

	$T = 100$			$T = 200$		
θ	full	i.i.d.	NC	full	i.i.d.	NC
0	0.049	0.043	0.054	0.045	0.045	0.046
0.25	0.064	0.079	0.090	0.112	0.120	0.131
0.5	0.150	0.162	0.195	0.324	0.327	0.371
0.75	0.282	0.289	0.340	0.572	0.582	0.641
1	0.380	0.415	0.505	0.781	0.790	0.857
1.25	0.525	0.523	0.654	0.894	0.903	0.938
1.5	0.610	0.644	0.751	0.958	0.942	0.969

Note: The nominal size is 5% and 1,000 replications are used. The column "full" refers to the tests using the correction $\hat{\psi}$ which allows for non-normal, conditionally heteroskesdatic and serially correlated errors, as defined by (8); 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 S.8: Size and power of the sup $LR^*_{1,T}(n_a = 1), UD\max LR_{1,T}$ and $CUSQ$ tests in the case of i.i.d. normal errors
(DGP: $y_t = e_t$; $e_t \sim i.i.d.\ N(0, 1 + \theta 1(t > [.5T]))$, $\varepsilon = 0.15$)

	$T = 100$			$T = 200$		
θ	sup $LR^*_{1,T}$	UDmax	$CUSQ$	sup $LR^*_{1,T}$	UDmax	$CUSQ$
0	0.049	0.051	0.030	0.045	0.044	0.029
0.25	0.064	0.064	0.059	0.112	0.108	0.116
0.5	0.150	0.136	0.142	0.324	0.302	0.351
0.75	0.282	0.259	0.268	0.572	0.554	0.613
1	0.380	0.356	0.391	0.781	0.762	0.808
1.25	0.525	0.497	0.521	0.894	0.889	0.918
1.5	0.610	0.588	0.599	0.958	0.951	0.965

Table S.9: Size of the sup $LR^*_{4,T}(m_a, n_a)$ and $UD\max LR_{4,T}$ tests in the case of i.i.d. normal errors (DGP: $y_t = e_t$, $e_t \sim i.i.d.\ N(0, 1)$)

	T=100			
ε	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax
0.2	0.039	0.043	0.042	0.042
0.15	0.040	0.035	0.043	0.039
0.1	0.041	0.043	0.042	0.042
	T=200			
ε	$m_a = n_a = 1$	$m_a = 1, n_a = 2$	$m_a = 2, n_a = 1$	UDmax
0.2	0.040	0.036	0.040	0.044
0.15	0.037	0.037	0.042	0.044
0.1	0.040	0.041	0.038	0.039

Table S.10: Finite sample performance of the split-sample procedure to select the number of breaks in coefficients and variance (DGP: $y_t = \mu_1 + \mu_2 1(t > T^c) + e_t$, $e_t \sim i.i.d.\ N(0, 1 + \theta 1(t > T^v))$, $\varepsilon = 0.15$, $T = 200$).

	$m = n = 0$	$m = n = 1$ $T^c = [.5T], T^v = [.7T]$				$m = n = 1$ $T^c = [.25T], T^v = [.75T]$		
		$\mu_2 = \theta = 1$	$\mu_2 = 1, \theta = 3$	$\mu_2 = 1, \theta = 5$	$\mu_2 = \theta = 2$	$\mu_2 = \theta = 1$	$\mu_2 = \theta = 2$	$\mu_2 = 1, \theta = 3$
$prob(m = 0, n = 0)$	**0.930**	0.002	0.000	0.000	0.000	0.014	0.000	0.015
$prob(m = 0, n = 1)$	0.022	0.000	0.009	0.014	0.000	0.001	0.000	0.006
$prob(m = 0, n = 2)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$prob(m = 1, n = 0)$	0.030	0.313	0.001	0.001	0.018	0.571	0.186	0.031
$prob(m = 1, n = 1)$	0.013	**0.618**	**0.880**	**0.856**	**0.922**	**0.331**	**0.714**	**0.822**
$prob(m = 1, n = 2)$	0.001	0.014	0.050	0.076	0.016	0.032	0.037	0.072
$prob(m = 2, n = 0)$	0.001	0.006	0.000	0.000	0.000	0.006	0.000	0.000
$prob(m = 2, n = 1)$	0.002	0.040	0.054	0.045	0.039	0.042	0.058	0.051
$prob(m = 2, n = 2)$	0.001	0.007	0.006	0.008	0.005	0.003	0.005	0.003
$prob(\bar{K} = 0)$	0.930	0.002	0.000	0.000	0.000	0.014	0.000	0.015
$prob(\bar{K} = 1)$	0.064	0.827	0.327	0.125	0.496	0.679	0.226	0.059
$prob(K = 2)$	0.006	0.171	0.673	0.875	0.504	0.307	0.774	0.926

	$m = n = 1$ $T^c = T^v = [.5T]$		$m = 1, n = 0$ $T^c = [.5T]$			$m = 0, n = 1$ $T^v = [.5T]$		
	$\mu_2 = \theta = 1$	$\mu_2 = 1, \theta = 3$	$\mu_2 = 1$	$\mu_2 = 2$	$\mu_2 = 3$	$\theta = 1$	$\theta = 2$	$\theta = 3$
$prob(m = 0, n = 0)$	0.000	0.000	0.000	0.000	0.000	0.379	0.020	0.000
$prob(m = 0, n = 1)$	0.001	0.009	0.000	0.000	0.000	**0.518**	**0.883**	**0.907**
$prob(m = 0, n = 2)$	0.000	0.000	0.000	0.000	0.000	0.002	0.005	0.007
$prob(m = 1, n = 0)$	0.080	0.000	**0.907**	**0.908**	**0.916**	0.007	0.000	0.001
$prob(m = 1, n = 1)$	**0.887**	**0.951**	0.056	0.061	0.048	0.080	0.073	0.069
$prob(m = 1, n = 2)$	0.004	0.010	0.005	0.008	0.004	0.001	0.003	0.006
$prob(m = 2, n = 0)$	0.000	0.000	0.024	0.016	0.017	0.001	0.001	0.000
$prob(m = 2, n = 1)$	0.023	0.022	0.004	0.006	0.011	0.009	0.013	0.010
$prob(m = 2, n = 2)$	0.005	0.008	0.004	0.001	0.004	0.003	0.002	0.000
$prob(\bar{K} = 0)$	0.000	0.000	0.000	0.000	0.000	0.379	0.020	0.000
$prob(\bar{K} = 1)$	0.962	0.951	0.960	0.962	0.958	0.588	0.940	0.966
$prob(K = 2)$	0.038	0.049	0.040	0.038	0.042	0.033	0.040	0.034

Note: $prob(m = j, n = i)$ represents the probability of choosing j breaks in mean and i breaks in variance, and $prob(\bar{K} = j)$ denotes the probability of selecting j total breaks in either mean or variance. The upper bound for the total number of breaks is set to 2.

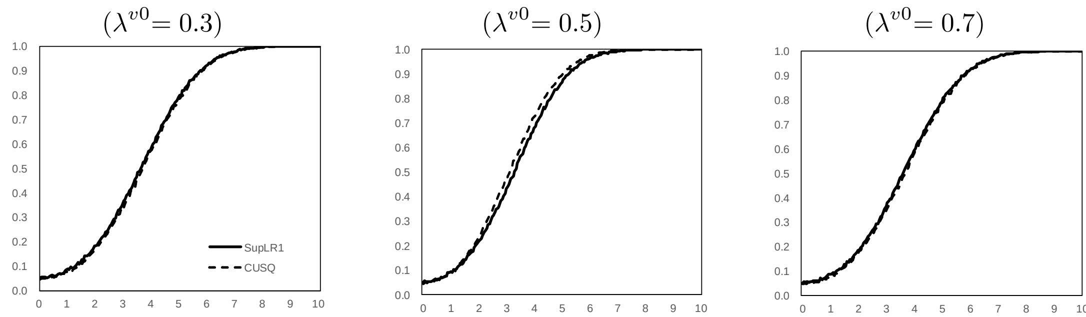

Figure S.1. Local asymptotic power functions of the sup $LR_{1,T}$ and CUSQ tests
($\lambda^{v0} = 0.3$, 0.5 and 0.7)

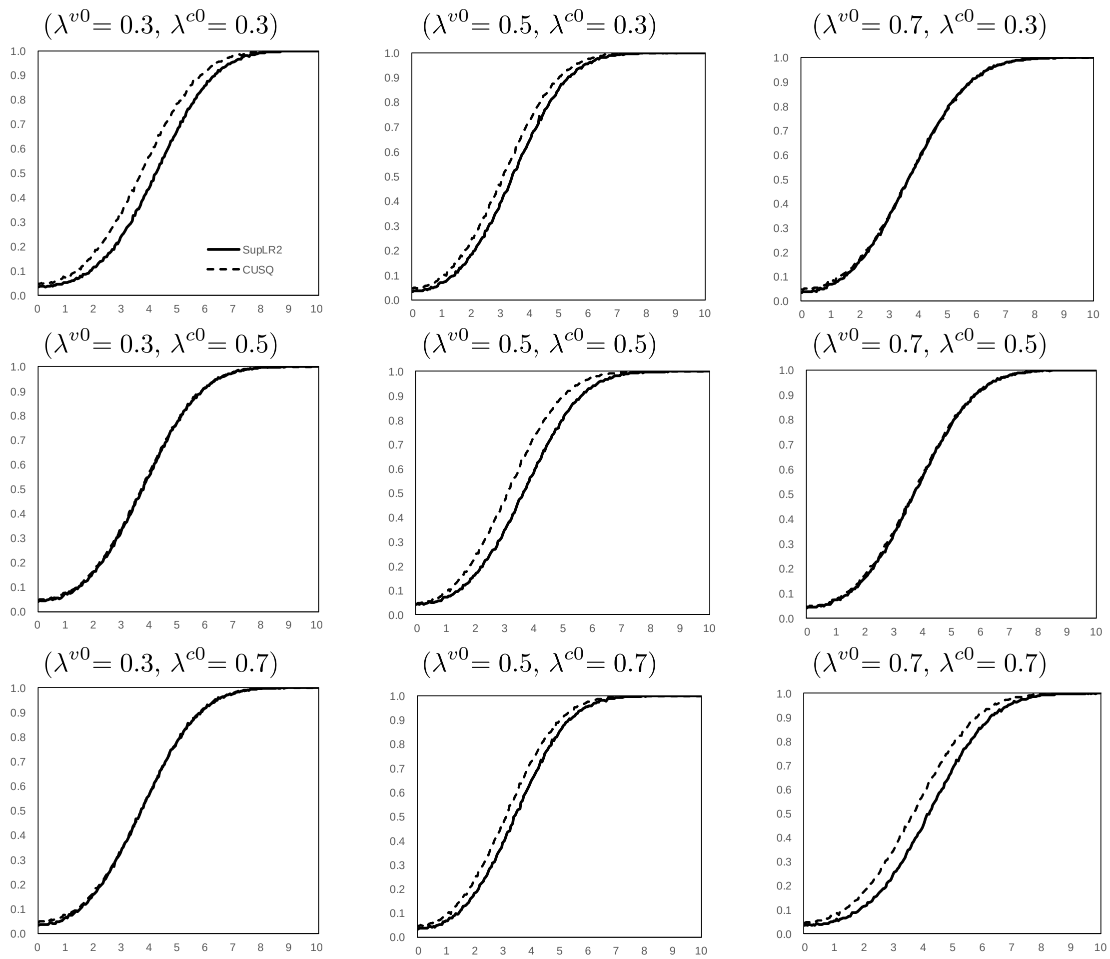

Figure S.2. Local asymptotic power functions of the sup $LR_{2,T}$ and CUSQ tests
($\lambda^{v0} = 0.3$, 0.5 and 0.7; $\lambda^{c0} = 0.3$, 0.5 and 0.7)