Economics Bulletin

Volume 46, Issue 3

Testing for structural breaks in the variance of estimators

Alaa Abi Morshed
Tilburg University

Abstract

In this paper, we shed light on breaks in the variance of estimators. To that end, we propose a sup-Wald structural break test for the variance of OLS estimators in a linear regression. The asymptotic distribution of our test is pivotal, although it is different from the standard asymptotic distributions of break point tests, at least for the heteroskedastic case. To obtain relevant critical values, we employ bootstrap methods . Additionally, a simulation study shows that our tests enjoy satisfactory empirical size and power properties.

Citation: Alaa Abi Morshed, (2026) "Testing for structural breaks in the variance of estimators", *Economics Bulletin*, Volume 46, Issue 3, pages 1246-1265
Contact: Alaa Abi Morshed - alaa.abi.morshed@gmail.com.
Submitted: September 21, 2024. **Published:** September 30, 2026.

1 Introduction

There is a plethora of literature related to testing for instability in parameters and error variance of regression models - see inter alia Andrews (1993), Inclan and Tiao (1994), and Bai and Perron (1998).

In this paper, we test for breaks in the variance of estimators because the latter captures all second-moment breaks in a regression setting. Therefore, we shed light on this source of breaks and study theoretically when does the distribution of the test differ from existing tests. Using the variance of estimators enhances power against a broader set of alternatives. Moreover, the test is particularly useful when the researcher is agnostic about the source of instability in the estimator's precision and seeks to avoid inflating Type 1 error when running separate tests for breaks in second moments.

We are the first to study testing for structural breaks in the variance of estimators. Additionally, our paper can be seen as an extension to the tests for breaks in variance literature, which is mainly focused on breaks in variance of the squared error terms see - Inclan and Tiao (1994) who propose a cusum test for breaks in the variance of error terms. We focus on linear models with one break which can occur in the second moments of regressors, squared error terms, or in the product of second moments of regressors and squared error terms (in case of conditional heteroskedasticity), at the same time. Breaks in the latter are new to the existing literature of structural breaks. In particular, we propose sup-Wald tests. In what follows we give illustrative examples to highlight the innovativeness of our test.

1.1 Illustrative Example

The object that we are dealing with is as follows:

$Var(\hat{\theta}) = (E[X_t X_t])^{-1} E(\epsilon_t^2 X_t X_t')(E[X_t X_t])^{-1}$

Breaks can occur in three components:

1.Error Variance: $E(\epsilon_t^2)$

2.Regressor Second Moments: $E[X_t X_t']$

3.Their interaction: $E(\epsilon_t^2 X_t X_t')$

First Example: Breaks arising from the joint distribution of regressors and conditional

error variance under conditional heteroskedasticity.

Let $\epsilon_t = \sigma(X_t)u_t$, $u_t \sim iid(0,1)$

Then, $E(\epsilon_t^2 X_t X_t') = E(\sigma(X_t)^2 X_t X_t')$

Suppose $\sigma(X_t)^2 = X_t^2$, then $E(\epsilon_t^2 X_t X_t')= E(X_t^4)$

Now even if, $Var(\epsilon_t)$ is stable unconditionally and $E(X_t^2)$ stable, a break in higher moments of X_t (for example in kurtosis, i.e., $E(X_t^4)$), changes $E(\epsilon_t^2 X_t X_t')$, which implies that the variance of estimator breaks, not the standard variance or covariance break tests. □

Second Example: Same coefficient, no error variance break, but estimator variance breaks.

Consider: $y_t = \beta X_t + \epsilon_t$, $\beta = 1$

Case A (no break in errors): $\epsilon \sim iid(0,1)$

Breaks occurs in variance of regressors: $X_t \sim \begin{cases} N(0,1), t < T/2 \\ N(0,4), t > T/2 \end{cases}$

First, β is constant: no break is detected by breaks in mean tests. Also, no break in error variance, so no break is returned by tests for breaks in error variance. Additionally, no break in the covariance structure. But variance of the estimator changes because $E(X_t^2)$ changes. □

On the related literature, Bai and Perron (1998) provide a comprehensive framework for testing and estimating multiple structural breaks in the coefficients of a linear regression model. The variance of the errors is allowed to be different across segments, but it is treated as a nuisance parameter. The focus is entirely on changes in the conditional mean. Perron, Yamamoto and Zhou (2020) extend the framework to jointly test for breaks in both the coefficients and the error variance. Their test is powerful when breaks occur in either or both components. My test is different from these. I do not test for breaks in mean at all. Instead, I focus solely on the variance of the estimates. This is a distinct object that can change even when there are no breaks in mean or variance of error terms (e.g., due to a break in regressors). Furthermore, as shown in the example, in the presence of heteroskedasticity, a break in variance of estimates can occur without a break in variance of error terms, but due to a break in the heteroskedasticity function itself, which the test of Perron, Yamamoto and Zhou (2020) would not detect.

Aue et al. (2009) proposes tests for detecting changes in the variance-covariance matrix of a time series itself. In my context, that would be the vector of regressors, so their test is directly applicable to detecting breaks in the variance of regressors. However, my test is more general. A break in the matrix of regressors is one potential source of a break in the variance of estimate, but not the only one. My test also captures breaks in the variance of error terms and breaks in the conditional heteroskedasticity function ($\mathbb{E}[\epsilon_t^2 X_t X_t']$) that are not captured by a break in variance of X_t alone. My test therefore covers a wider array of second-moment instabilities that affect inference.

Inclan and Tiao (1994), is a classic reference for testing for a break in the variance of error terms in a regression context. Dette and Wied (2016) provide tests for breaks in the variance of errors in a more general time series setting. My test encompasses a break in variance of error terms as a special case (under homoskedasticity). More importantly, my test remains valid and powerful in more complex scenarios where a break occurs in the product $\mathbb{E}[\epsilon_t^2 X_t X_t']$, even if variance of error terms remains constant. This scenario is new to the structural break literature and is a key innovation of my paper.

We establish the asymptotic null distribution of our test and find that the latter is pivotal but different from the standard distribution of break point tests, for the conditional heteroskedastic case. This gives rise to a new empirical process which is new to the structural break literature. Therefore, we bootstrap the corresponding distribution and obtain relevant critical values. Simulations using the latter show that the test enjoys satisfactory size and power properties.

This paper is organized as follows. In Section 2, we put out the model, test statistic, and assumptions used. Section 3 states the functional central limit theorem (FCLT) and the asymptotic null distributions of the tests, under homoskedasticity and heteroskedasticity, respectively. Section 4 is devoted for a simulation study. Proofs of lemmas and theorems are relegated to Appendix A. Lastly, we mention notational conventions that are used throughout the paper: *vech* is the operator that stacks the columns on and below the diagonal of a symmetric p × p matrix as a vector with p(p+1)/2 components. Also $\otimes$ is the widely known kronecker product. $\Rightarrow$ denotes weak convergence under the Skorohod topology. And $\sum_{1\lambda} = \sum_{t=1}^{[T\lambda]}$ and $\sum_{2\lambda} = \sum_{t=[T\lambda]+1}^{T}$. Additionally, AVar = $\lim_{T\to\infty} Var$

2 The model and assumptions

The true model takes the general form:

$$y_t = X_t'\theta + \epsilon_t, \tag{1}$$

where X_t is a $p \times 1$ vector of (exogenous) regressors that includes an intercept and possibly lagged dependent variables and ϵ_t are scalar errors with variance, v_ϵ. $\hat{\theta}$ is a $p \times 1$ vector of estimates whose variance is subject to structural breaks across the sample.

The conditional var-covar matrix of $\hat{\theta}$ under homoskedasticity is:

$$Var(\hat{\theta})_{i\lambda} = v_{\epsilon,i\lambda} \left(T^{-1} \sum_{i\lambda} X_t X_t' \right)^{-1}, (i = 1, 2)$$

and under heteroskedasticity,

$$Var(\hat{\theta})_{i\lambda} = \left(T^{-1} \sum_{i\lambda} X_t X_t' \right)^{-1} \left(T^{-1} \sum_{i\lambda} E(\epsilon_t^2 X_t X_t') \right) \left(T^{-1} \sum_{i\lambda} X_t X_t' \right)^{-1}, (i = 1, 2)$$

Let $V^{vec}_{\hat{\theta}_{1\lambda}} = vech[Var(\hat{\theta})_{1\lambda}]$ and $V^{vec}_{\hat{\theta}_{2\lambda}} = vech[Var(\hat{\theta})_{2\lambda}]$.

where

$$V^{vec}_{\hat{\theta}_t} = \begin{cases} V^{vec}_{\hat{\theta}_{1\lambda}} 1[t \leq T_0^V] \\ V^{vec}_{\hat{\theta}_{2\lambda}} 1[t > T_0^V] \end{cases} \tag{2}$$

and $T_0^V = [T\lambda_V]$, $0 < \lambda_V < 1$ are the true break dates and fractions.

We are interested in tests of the following null and alternative hypothesis.

$$H_0 : V^{vec}_{\hat{\theta}_t} = V^{vec}_{\hat{\theta}} \qquad vs. \qquad H_A : V^{vec}_{\hat{\theta}_{1\lambda}} \neq V^{vec}_{\hat{\theta}_{2\lambda}}$$

The corresponding sup-Wald test for breaks in the conditional var-covariance matrix is as follows:

$$W_T = \sup_{\lambda \in [\epsilon, 1-\epsilon]} W_T(\lambda)$$

where

$$W_T(\lambda) = T \left(\hat{V}^{vec}_{\hat{\theta}_{1\lambda}} - \hat{V}^{vec}_{\hat{\theta}_{2\lambda}} \right)' \widehat{V}_\lambda^{-1} \left(\hat{V}^{vec}_{\hat{\theta}_{1\lambda}} - \hat{V}^{vec}_{\hat{\theta}_{2\lambda}} \right)$$

and $\hat{V}^{vec}_{\hat{\theta}_{1\lambda}} = vech[\widehat{Var}(\widehat{\theta})_{1\lambda}]$, $\hat{V}^{vec}_{\hat{\theta}_{2\lambda}} = vech[\widehat{Var}(\widehat{\theta})_{2\lambda}]$, $\widehat{Var}(\widehat{\theta})_{1\lambda}$, $\widehat{Var}(\widehat{\theta})_{2\lambda}$ are the sample var-covariance matrices of $\hat{\theta}$ in (1) in subsamples $\{1, \ldots, T_{1\lambda}\}$ and $\{T_{1\lambda}+1, \ldots, T\}$, respectively where $T_{1\lambda}$ stands for $[T\lambda]$. And $\widehat{\Omega}_\lambda$ is a consistent estimator of $AVar \left(T^{1/2} \left(\hat{V}^{vec}_{\hat{\theta}_{1\lambda}} - \hat{V}^{vec}_{\hat{\theta}_{2\lambda}} \right) \right)$

under H_0. Additionally, ϵ is the standard trimming parameter in the break point literature and $V_{\hat{\theta}}^{vec} = vech[Var(\hat{\theta})_0]$.

The long run variance, $\hat{\Omega}$, takes the following form:

Let $\hat{Z}_t = \text{vech}(\hat{\epsilon}_t^2 X_t X_t')$, then

$\hat{\Omega} = \sum_{k=-K_T}^{K_T} w\left(\frac{k}{K_T}\right)\hat{\Gamma}_k$ *where* $\hat{\Gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T}(\hat{Z}_t - \bar{\hat{Z}})(\hat{Z}_{t-k} - \bar{\hat{Z}})$ and w is a kernel.

The sample conditional variance-covariance matrix of $\hat{\theta}$ under homoskedasticity is:

$$\begin{aligned}&\widehat{Var}(\widehat{\theta})_{i\lambda} = \widehat{v}_{\epsilon,i\lambda}\left(T^{-1}\textstyle\sum_{i\lambda} X_t X_t'\right)^{-1}, \widehat{v}_{\epsilon,i\lambda} = T^{-1}\textstyle\sum_{i\lambda}\widehat{\epsilon}_t^2, (i=1,2).\\ &\widehat{\epsilon}_t = y_t - X_t'\widehat{\theta}.\end{aligned} \tag{3}$$

and under heteroskedasticity,

$$\widehat{Var}(\widehat{\theta})_{i\lambda} = \left(T^{-1}\textstyle\sum_{i\lambda} X_t X_t'\right)^{-1}\left(T^{-1}\textstyle\sum_{i\lambda}\widehat{\epsilon}_t^2 X_t X_t'\right)\left(T^{-1}\textstyle\sum_{i\lambda} X_t X_t'\right)^{-1}, (i=1,2). \tag{4}$$

Specifically, we make the following assumptions:

Assumption 1. *$E(\epsilon_t) = 0$, $E(x_t\epsilon_t) = 0$ and the matrix $\sum_{t=1}^{[T\lambda]} X_t X_t'$ is invertible for all values of $\lambda \in [\epsilon, 1-\epsilon]$.*

Assumption 2. *$T^{-1}\sum_{t=1}^{[T\lambda]} X_t X_t' \xrightarrow{p} \lambda Q$, a positive definite (pd) matrix of constants.*

Assumption 3. *Define $\mathcal{F}_t$=σ-field $\{,,, X_{t-1}X_{t-1}', X_t X_t', \epsilon_{t-1}^2, \epsilon_t^2, \epsilon_{t-1}^2 X_{t-1}X_{t-1}', \epsilon_t^2 X_t X_t'\}$ and let $v_{x\epsilon} = E(X_t X_t'\epsilon_t^2)$, under the null.*

If $\{X_t X_t'\epsilon_t^2\}$ is weakly stationary, then

a) $\{vech(X_t X_t'\epsilon_t^2 - v_{x\epsilon}), \mathcal{F}_t\}$ forms a strongly mixing (α-mixing) with size $\frac{-4(4+\delta)}{\delta}$ where δ is some positive constant,

b) $\sup_t \|vech(X_t X_t'\epsilon_t^2 - v_{x\epsilon})\|_{2(4+\delta)} < C_1$ and

c) let $S_T(m) = \sum_{t=1+m}^{T+m}\eta_t$ where $\eta_t = vech(X_t X_t'\epsilon_t^2 - v_{x\epsilon})$ and let $S_{iT}(m)$ be the ith component of $S_T(m)$, hereafter $S_T = S_T(0)$. If $\{x_t x_t'\epsilon_t^2\}$ is not weakly stationary within each segment, we assume that (a)-(c) holds and, in addition, that there exists a positive definite matrix, $M_{x\epsilon} = [w_{i,j}]$, such that for $\forall$i,j=1,2,..,p(p+1)/2, we have uniformly in m, $|T^{-1}E\left(S_{iT}(m)S_{jt}(m)\right) - w_{i,j}| \leq CT^{-\psi}$ for some C, $\psi > 0$ and C is independent of m.

Assumption 4. *Under conditional homoskedasticity, then Assumption 3 holds with* $X_t X_t' \epsilon_t^2 - v_{x\epsilon}$ *replaced with* $\epsilon_t^2 - v_\epsilon$.

Assumption 1 presents sufficient conditions for the identification of OLS estimates and states that the regressors are exogenous.[1] Assumption 2 allows for shrinking breaks in the second moment of the regressors. This assumption is used to arrive at the asymptotic distributions of the tests under the null. Although this assumption may sound restrictive, it allows for breaks in the second moment of the regressors in finite samples, however, in the limit these breaks disappear. Additionally, unit root regressors are ruled out.

Assumptions 3 and 4 determine the dependence structure of the processes $\{\epsilon_t^2\}$ and $\{X_t X_t' \epsilon_t^2\}$. They imply that the processes are short memory but can exhibit substantial autocorrelation and conditional heteroskedasticity. These assumptions are very similar to the ones in Qu and Perron (2007). Moreover, Assumptions 3 and 4 are needed to obtain a strong invariance principle for the homoskedastic and heteroskedastic cases respectively - see Lemma 1 in the section below.

3 FCLT and Asymptotic Null Distribution

In this section, we state in Lemma 1 that a FCLT holds for the processes at hand. Lastly, we establish the asymptotic null distributions of our test, under (conditional) homoskedasticity and heteroskedasticity, respectively in Theorems 1 and 2.

Lemma 1.

Under conditional homoskedasticity and Assumption 4
we have a) $Var\left(T^{-1/2}\sum_{t=1}^{[T\lambda]}\epsilon_t^2\right) = \lambda\gamma^2$, *where* γ^2 *is the variance of squared error terms*[2] *and there exists a Brownian Motion* $(\mathcal{B}(t)_{t>0})$ *with variance* γ^2 *such that* $\sum_{i=1}^{t}(\epsilon_t^2 - v_\epsilon) - \mathcal{B}(t) = O_{a.s}(t^{1/2-\kappa})$ *for* $\kappa > 0$. *b) (FCLT)* $T^{-1/2}\sum_{t=1}^{[T\lambda]}(\epsilon_t^2 - v_\epsilon) \Rightarrow \gamma\mathcal{B}_1(\lambda)$ *where* $\mathcal{B}_1(\lambda)$ *is a scalar independent brownian motion.*

Under conditional heteroskedasticity and Assumption 3
we have a) $AVar\left(T^{-1/2}\sum_{t=1}^{[T\lambda]} vech\big(X_t X_t'\epsilon_t^2\big)\right) = \lambda\Omega$, *where* Ω *is the corresponding* $p(p+1)/2 \times p(p+1)/2$ *long run covariance matrix and there exists a vector of Brownian Motions* $(\mathcal{B}(t)_{t>0})$ *with variance* Ω *such that* $\sum_{i=1}^{t} vech(X_t X_t'\epsilon_t^2 - v_{x\epsilon}) - \mathcal{B}(t) = O_{a.s}(t^{1/2-\kappa})$ *for* $\kappa > 0$. *b) (FCLT)* $T^{-1/2}\sum_{t=1}^{[T\lambda]} vech\left(X_t X_t'\epsilon_t^2 - v_{x\epsilon}\right) \Rightarrow \Omega^{1/2}\mathcal{B}_{p(p+1)/2}(\lambda)$ *where* $\mathcal{B}_{p(p+1)/2}(\lambda)$

[1] If we instead allow lagged dependent variables, we assume in addition, that all the characteristic roots associated with the lag polynomials for the lagged dependent variables are inside the unit circle.

[2] The variance might include the covariances, depending whether we allow for autocorrelation or not.

is a vector of independent Brownian Motions and $\Omega^{1/2}$ *is the* $p(p+1)/2 \times p(p+1)/2$ *square root matrix of* Ω.

In Theorems 1 and 2 below, we establish the asymptotic null distribution of the test, under conditional homoskedastcity and heteroskedasticity, respectively.

Theorem 1. ***Under Homoskedasticity of error terms***[3]

Let the model be as in (1) *and* (2) *, and let Assumptions 1-2, and 4, and Lemma 1 hold. Then under* H_0, $W_T \Rightarrow \sup_{\lambda \in [\epsilon, 1-\epsilon]} \mathcal{G}_1(\lambda)$ *where* $\mathcal{G}_1(\lambda) = \frac{[\mathcal{B}_1(\lambda) - \lambda \mathcal{B}_1(1)]^2}{\lambda(1-\lambda)}$.[4]

Theorem 1 states the asymptotic distribution of our test is pivotal and is identical to the distribution obtained for conditional mean tests with one single regressor see-Andrews (1993) and Andrews and Ploberger (1994). Therefore, tabulated asymptotic critical values can be used.

Theorem 2. ***Under Heteroskedasticity of error terms***

Let the model be as in (1) *and* (2) *, and let Assumptions 1-3 and Lemma 1 hold. Then under* H_0, $W_T \Rightarrow \sup_{\lambda \in [\epsilon, 1-\epsilon]} \mathcal{G}_{(p)(p+1)/2}(\lambda)$

where $\mathcal{G}_{(p)(p+1)/2}(\lambda) = \frac{1}{\lambda^3(1-\lambda)^3} \left(J_{p(p+1)/2}(\lambda) \right)' \cdot \left(J_{p(p+1)/2}(\lambda) \right)$

and $J_{p(p+1)/2}(\lambda) = (1-\lambda)^2 \mathcal{B}_{p(p+1)/2}(\lambda) - \lambda^2 [\mathcal{B}_{p(p+1)/2}(1) - \mathcal{B}_{p(p+1)/2}(\lambda)]$.[5]

Theorem 2 states the asymptotic distribution of our test is pivotal though it is different from the standard break point tests' asymptotic distributions. In particular, and in comparison to the already existing distributions for one-break tests, a new empirical process arises.

Remark. The asymptotic distribution under conditional heteroskedasticity differs from the standard distribution for break point tests in the following aspects: First, its not a brownian bridge, but a new transformation of brownian motion. Second, weights are not linear but quadratic, and normalization is different. We revert in the next section, to bootstrap the asymptotic distribution of this new process to obtain correctly sized tests.

4 Simulations

We consider in our simulation study two data generating processes that correspond to breaks in conditional heteroskedasticity, homoskedasticity, respectively. We also study

[3]By assuming homoskedasticity, we mean by it, conditional homoskedasticity where variance of error terms not a function of observables, but this does not restrict or preclude changes in unconditional variances, for example breaks in the latter.

[4]Additionally, there is an implicit underlying assumption that there is no break in the conditional mean (i.e., θ), otherwise the asymptotic distribution of our test is not valid. This is a standard assumption for breaks in variance tests.

[5]Additionally, there is an implicit underlying assumption that there is no break in the conditional mean (i.e., θ), otherwise the asymptotic distribution of our test is not valid. This is a standard assumption for breaks in variance tests.

the impact of a change in the magnitude of the break and its location on the power of the test.

We employ the fixed regressor Wild bootstrap for the data generating process under conditional heteroskedasticity and the fixed regressor IID bootstrap for the data generating process under conditional homoskedasticity.

4.1 Data Generating Processes

DGP1: Breaks in Conditional Heteroskedasticity

Null

$y_t = \alpha + X_t\beta + \epsilon_t$ where $X_t \overset{\text{iid}}{\sim} \mathcal{N}(1,1)$ and $\epsilon_t \overset{\text{id}}{\sim} \mathcal{N}(0, \sigma_\epsilon(X_t))$.
where $\sigma_\epsilon(X_t) = 0.4(X_t^2)$ and α=β=1.

Alternative

Break in second moment of regressor where $\Delta\sigma_x = \delta_x$ leads to a break in σ_ϵ.

$$E(x_t^2) = \begin{cases} \sigma_{1x} 1[t \leq T_0^x] \\ \sigma_{2x} 1[t > T_0^x]. \end{cases}$$

where $T_0^x = 0.5$.

DGP2: Breaks in Conditional Homoskedasticity

Null

$y_t = \alpha + x_t\beta + \epsilon_t$ where $x_t \overset{\text{iid}}{\sim} \mathcal{N}(1,1)$ and $\epsilon_t \overset{\text{id}}{\sim} \mathcal{N}(0, \sigma_\epsilon)$ and α=β=1.

Alternative

Same as above, but here we have a break in $Var(\epsilon_t)$, $\Delta\sigma_\epsilon = \delta_\epsilon$, at a certain point.

$$Var(\epsilon_t) = \begin{cases} \sigma_{1\epsilon} \mathbf{1}[t \leq T_0^E] \\ \sigma_{2\epsilon} \mathbf{1}[t > T_0^E] \end{cases}$$

where $T_0^\epsilon = 0.5$.

The results in Tables 1-5 below are generated employing 1000 Monte Carlo repetitions[6], 399 Bootstraps, a trimming parameter equal to 0.15, and significance level equal

[6]We stick to this number of Monte Carlo iterations since the desired empirical size/power is already reached.

to 0.05.

With respect to the size of the test in Table 1, we notice that the size for DGP's 1 and 2 is close to the nominal size across all sample sizes. Additionally, we study the impact of a change in the magnitude of a break in either σ_x or σ_ϵ as in DGPs 1 and 2, on the power properties of the test. Tables 2 and 3 show that the power of the test approaches 1 as the break magnitude and sample size increase, in line with expectations.

In Tables 4 and 5, we study the impact of a change in the true break date on the power properties of the test. In particular, we study the impact of earlier (T^0=0.3) and later breaks (T^0=0.7) in the sample. Comparing Tables 2, 4 and 5 for the test under conditional heteroskedasticity, we find that power is a little bit lower for small samples when the true break is earlier in the sample and a bit higher when the true break date is later in the sample. This effect disappears for larger samples and magnitudes of breaks.

Table 1: Empirical Size for DGP1 and DGP2

Empirical Size	T200	T500	T1000
DGP_1	0.078	0.050	0.054
DGP_2	0.048	0.046	0.04

[a] 1000 Monte Carlo repetitions, T200 until T1000 are the sample sizes, trimming parameter, ϵ, is 0.15.

Table 2: Empirical Power for DGP1, T_x^0=0.5

Magnitude of Break	T200	T500	T1000
$\delta_x = 1$	0.09	0.58	0.97
$\delta_x = 2$	0.25	0.87	1
$\delta_x = 3$	0.40	0.95	1
$\delta_x = 4$	0.47	0.96	1
$\delta_x = 5$	0.55	0.98	1
$\delta_x = 6$	0.58	0.99	1
$\delta_x = 7$	0.60	0.98	1
$\delta_x = 8$	0.65	0.99	1

[a] 1000 Monte Carlo repetitions, δ_x corresponds to magnitude of break in σ_x, T200 until T1000 are the sample sizes, trimming parameter, ϵ, is 0.15, $T_x^0 = 0.5$.

Table 3: Empirical Power for DGP2, T_x^0=0.5

Magnitude of Break	T200	T500	T1000
$\delta_\epsilon = 1$	0.80	1	1
$\delta_\epsilon = 2$	0.970	1	1
$\delta_\epsilon = 3$	0.992	1	1
$\delta_\epsilon = 4$	0.995	1	1
$\delta_\epsilon = 5$	0.996	1	1
$\delta_\epsilon = 6$	0.997	1	1
$\delta_\epsilon = 7$	0.998	1	1
$\delta_\epsilon = 8$	1	1	1

[a] 1000 Monte Carlo repetitions, δ_ϵ corresponds to magnitude of break in σ_e, T200 until T1000 are the sample sizes, trimming parameter, ϵ, is 0.15, $T_\epsilon^0 = 0.5$.

Table 4: Empirical Power for DGP1, T_x^0=0.3

Magnitude of Break	T200	T500	T1000
$\delta_x = 1$	0.04	0.35	0.82
$\delta_x = 2$	0.17	0.74	0.96
$\delta_x = 3$	0.28	0.86	0.99
$\delta_x = 4$	0.31	0.91	1
$\delta_x = 5$	0.35	0.92	1
$\delta_x = 6$	0.36	0.94	1
$\delta_x = 7$	0.40	0.95	1
$\delta_x = 8$	0.43	0.96	1

[a] 1000 Monte Carlo repetitions, δ_x corresponds to magnitude of break in σ_x, T200 until T1000 are the sample sizes, trimming parameter, ϵ, is 0.15, $T_x^0 = 0.3$.

Table 5: Empirical Power for DGP1, T_x^0=0.7

Magnitude of Break	T200	T500	T1000
$\delta_x = 1$	0.19	0.82	0.99
$\delta_x = 2$	0.47	0.97	1
$\delta_x = 3$	0.59	0.99	1
$\delta_x = 4$	0.67	1	1
$\delta_x = 5$	0.68	1	1
$\delta_x = 6$	0.73	1	1
$\delta_x = 7$	0.74	1	1
$\delta_x = 8$	0.76	1	1

[a] 1000 Monte Carlo repetitions, δ_x corresponds to magnitude of break in σ_x, T200 until T1000 are the sample sizes, trimming parameter, ϵ, is 0.15, $T_x^0 = 0.7$.

5 Conclusion

Testing for breaks in the variance of estimators captures all second moments in a regression setting. In particular, breaks in the product of regressors and squared error terms (i.e., conditional heteroskedasticity), which is new to the existing break point literature. Thus, we propose a formal testing procedure to test for structural breaks in the variance of OLS estimators. We establish the asymptotic null distributions which are pivotal but different from the standard distributions for break point tests when errors are heteroskedastic. Moreover, a simulation study shows that our test enjoys satisfactory size and power properties.

On potential empirical applications, our test can be applied to study the stability of the Philips Curve. The main motivation is that the literature focuses on breaks in coefficients and error variance, but even if both are stable, changes in the joint second moments can break inference as in the case of our test. For example, in post-2008 or COVID period, volatility depends on macro conditions and regressors have become more dispersed. In this case, our test detects instability in precision of Philips curve estimates, not just the slope.

6 Appendix A

Proof of Lemma 1:

Given assumptions 3 and 4, the results in Lemma 1 follow from Theorem 1 in Eberlein (1986), when the following two expressions have been shown: **A.1**-$\|E(S_T(m)/\mathcal{F}_m)\|_2 \leq C$ for C independent of m and $\mathcal{F}_m$ as defined above. **A.2**- $\forall$ i, j=1,2,.., p(p+1)/2, we have

uniformly in m, $T^{-1}E|E\left(S_{iT}(m)S_{jt}(m)/\mathcal{F}_m\right) - E\left(S_{iT}(m)S_{jt}(m)\right)| = O(T^{-\theta})$ for some $\theta > 0$.

We begin by showing **A.1**.

By Minkowski's inequality, Lemma 3.5 in McLeish (1975) and Assumption 3c),

$$\begin{aligned}
||E(S_T(m)||_2 &\leq ||E(\eta_{m+1}/F_m)||_2 + ||E(\eta_{m+2}/F_m)||_2 + ... + ||E(\eta_{m+T}/F_m)||_2 \\
&\leq 6\alpha_1^{1/2-1/2(4+\delta)}||\eta_{m+1}||_{2(4+\delta)} + .. + 6\alpha_T^{1/2-1/2(4+\delta)}||\eta_{m+T}||_{2(4+\delta)} \\
&\leq 6\sup_t ||\eta_t||_{2(4+\delta)} \sum_{j=1}^{\infty} \alpha_j^{1/2-1/2(4+\delta)} \\
&\leq 6C_1 \sum_{j=1}^{\infty} \alpha_j^{1/2-1/2(4+\delta)} \leq 6C_1 A \leq C.
\end{aligned}$$

The final inequality follows from the bound in Assumption 3 b)-c) and because given Assumption 3 a), $\sum_{j=1}^{\infty} \alpha_j^{1/2-1/2(4+\delta)} = A \leq \infty$.

It remains to prove **A.2**.

Recall that $S_{iT}(m) = \sum_{k=m+1}^{m+T} \eta_k$ and $S_{iT}(m+a) = \sum_{m+a+1}^{m+a+T} \eta_k$. The left hand side (LHS) of **A.2** can be rewritten as follows:

$$\begin{aligned}
&E|E(T^{-1}S_{iT}(m+a)S_{jT}(m+a)/F_m) - E(T^{-1}S_{iT}(m+a)S_{jT}(m+a)) \\
&-E(T^{-1}S_{iT}(m+a)S_{jT}(m+a)) - E(T^{-1}S_{iT}(m)S_{jT}(m)) \\
&+E(T^{-1}S_{iT}(m+a)S_{jT}(m+a)) - E(T^{-1}S_{iT}(m)S_{jT}(m))| \\
&\leq \\
&E|E(T^{-1}S_{iT}(m+a)S_{jT}(m+a)/F_m) - E(T^{-1}S_{iT}(m+a)S_{jT}(m+a))| \\
&+2E|T^{-1}S_{iT}(m+a)S_{jT}(m+a) - T^{-1}S_{iT}(m)S_{jT}(m)| \quad \textbf{(A.3)}.
\end{aligned}$$

We shall show that the first term on the right hand side (RHS) of the inequality in (A.3) is $O(T^{-\theta_1})$ and the second term is $O(T^{-\theta_2})$, $\theta_1, \theta_2 > 0$. Equation(A.2) then follows by setting $\theta = Min(\theta_1, \theta_2)$.

As for the first term on the RHS of (A.3), from McLeish mixing inequality (1975a, Lemma 3.5), Cauchy-Schwartz and Minkowski inequality, we have

$$T^{-1}E|E(T^{-1}S_{iT}(m+a)S_{jT}(m+a)/F_m) - E(T^{-1}S_{iT}(m+a)S_{jT}(m+a))|$$

$$\leq 6T^{-1}\alpha_a^{1-1/(2+\delta)}||S_{iT}(m+a)S_{jT}(m+a)||_{2+\delta}$$

$$\leq 6T^{-1}\alpha_a^{\frac{1+\delta}{2+\delta}}||S_{iT}(m+a)||_{2(2+\delta)}||S_{jT}(m+a)||_{2(2+\delta)}$$

$$\leq 6T^{-1}T^{2}\alpha_a^{\frac{1+\delta}{2+\delta}} sup_t||\eta_t||_{2(2+\delta)}$$
$$\leq 6C_1 T\alpha_a^{\frac{1+\delta}{2+\delta}}.$$

where C_1 is the bound in Assumption 3 b)-c). Given the size conditions in Assumption 3 a) and for $a = T^{\phi}$, with $\frac{\delta(2+\delta)}{4(\delta+1)(4+\delta)} < \phi < 1$, $T\alpha_a^{\frac{1+\delta}{2+\delta}}$ will approach zero as T tends to ∞ at rate of $T^{-\theta_1}$ with $\theta_1 = \phi\frac{4(\delta+1)(4+\delta)}{\delta(2+\delta)} - 1$.

We now need to show that by setting $a = T^{\phi}$, with ϕ defined as previously, the second term on the RHS of A.3 is $O(T^{-\theta_2})$, $\theta_2 > 0$. First note that,

$$\begin{aligned}
&T^{-1}S_{iT}(m+a)S_{jT}(m+a) - T^{-1}S_{iT}(m)S_{jT}(m) \\
&= (T^{-1/2}S_{jT}(m+a) - T^{-1/2}S_{jT}(m))(T^{-1/2}S_{iT}(m+a) - T^{-1/2}S_{iT}(m)) \\
&\quad + T^{-1/2}S_{jT}(m+a)(T^{-1/2}S_{iT}(m+a) - T^{-1/2}S_{iT}(m)) \\
&\quad - T^{-1/2}S_{iT}(m+a)(T^{-1/2}S_{jT}(m+a) - T^{-1/2}S_{jT}(m)).
\end{aligned}$$

Thus by the triangular inequality and Cauchy-Schwartz inequality,

$$\begin{aligned}
&E|T^{-1}S_{iT}(m+a)S_{jT}(m+a) - T^{-1}S_{iT}(m)S_{jT}(m)| \\
&\quad \leq (E(T^{-1/2}S_{jT}(m+a) - T^{-1/2}S_{jT}(m))^2)^{1/2} \times (E(T^{-1/2}S_{iT}(m+a) - T^{-1/2}S_{iT}(m))^2)^{1/2} \\
&\qquad + (E(T^{-1/2}S_{jT}(m+a))^2)^{1/2}(E(T^{-1/2}S_{iT}(m+a) - T^{-1/2}S_{iT}(m))^2)^{1/2} \\
&\qquad + (E(T^{-1/2}S_{iT}(m+a))^2)^{1/2}(E(T^{-1/2}S_{jT}(m+a) - T^{-1/2}S_{jT}(m))^2)^{1/2} \\
&= O(a^{1/2}T^{-1/2}).
\end{aligned}$$

The above follows from Assumption 3 c) where $E(T^{-1/2}S_{iT}(m+a) - T^{-1/2}S_{iT}(m))^2 = O(aT^{-1})$ and $E(T^{-1/2}S_{iT}(m+a))^2 = O(1)$, uniformly in m, for a =o(T). Therefore, by setting $a = T^{\phi}$, we have that the second term in $A.3$ is $O(T^{-\theta_2})$ with $\theta_2 = \frac{1}{2}(1-\phi)$.

The above proof follows similar steps as in Corradi (1999) (1999, p. 651-653) for $\eta_t = vech(x_t x_t'\epsilon_t^2 - v_{x\epsilon})$ or $\eta_t = \epsilon_t^2 - v_\epsilon$ respectively. □

Proof of Theorem 1:

It follows from the homoskedasticity assumption that,

$$\sqrt{T}\hat{V}_{\hat{\theta}_{i\lambda}}^{vec} = \sqrt{T}\widehat{v}_{\epsilon,i\lambda} vech[(T^{-1}\textstyle\sum_{i\lambda} X_t X_t')^{-1}] \qquad ,(i=1,2).$$

For i=1 and making use of the consistency of $\widehat{v}_{\epsilon,1\lambda}$, (i.e., $\widehat{v}_{\epsilon,1\lambda} = \lambda v_\epsilon + \mathcal{O}_{p(1)}$), the following obtains:

$$\begin{aligned}\sqrt{T}\hat{V}^{vec}_{\hat{\theta}_{1\lambda}} &= \sqrt{T}\widehat{v}_{\epsilon,1\lambda} vech[(T^{-1}\textstyle\sum_{1\lambda} X_t X_t')^{-1}] = \sqrt{T}\left(\lambda v_\epsilon + \mathcal{O}_{p(1)}\right) vech[(T^{-1}\textstyle\sum_{1\lambda} X_t X_t')^{-1}] \\ &= \sqrt{T}\left(\widehat{v}_{\epsilon,1\lambda} - \lambda v_\epsilon\right) vech[(T^{-1}\textstyle\sum_{1\lambda} X_t X_t')^{-1}] + \mathcal{O}_{p(1)}{}^7 \\ &= \sqrt{T}\left(\frac{\sum_{1\lambda}\widehat{\epsilon}_t^2}{T} - \frac{\lfloor T\lambda\rfloor}{T} v_\epsilon\right) vech[(T^{-1}\textstyle\sum_{1\lambda} X_t X_t')^{-1}] + \mathcal{O}_{p(1)} \\ &= T^{-1/2}\textstyle\sum_{1\lambda}\left(\widehat{\epsilon}_t^2 - v_\epsilon\right) vech[(T^{-1}\textstyle\sum_{1\lambda} X_t X_t')^{-1}] + \mathcal{O}_{p(1)}\end{aligned} \tag{5}$$

In a similar fashion, for i=2 and making use of the consistency of $\widehat{v}_{\epsilon,2\lambda}$, (i.e., $\widehat{v}_{\epsilon,2\lambda} = (1-\lambda)v_\epsilon + \mathcal{O}_{p(1)}$), the following obtains:

$$\sqrt{T}\hat{V}^{vec}_{\hat{\theta}_{2\lambda}} = T^{-1/2}\textstyle\sum_{2\lambda}\left(\widehat{\epsilon}_t^2 - v_\epsilon\right) vech[(T^{-1}\textstyle\sum_{2\lambda} X_t X_t')^{-1}] + \mathcal{O}_{p(1)}$$

And from Assumption 4 and Lemma 1, it follows

$$T^{-1/2}\textstyle\sum_{t=1}^{[T\lambda]}\left(\hat{\epsilon}_t^2 - v_\epsilon\right) \Rightarrow \gamma\mathcal{B}_1(\lambda) \quad and \;\; T^{-1/2}\textstyle\sum_{t=[T\lambda]+1}^{T}\left(\hat{\epsilon}_t^2 - v_\epsilon\right) \Rightarrow \gamma[\mathcal{B}_1(1) - \mathcal{B}_1(\lambda)]$$

By Uniform Law of Large numbers, Assumption 2 and the continuous mapping theorem, it follows that:

$$vech\left[\left(T^{-1}\sum_{1\lambda} X_t X_t'\right)^{-1}\right] \xrightarrow{p} vech\left[(\lambda Q)^{-1}\right]$$

and

$$vech\left[\left(T^{-1}\sum_{2\lambda} X_t X_t'\right)^{-1}\right] \xrightarrow{p} vech\left[((1-\lambda) Q)^{-1}\right]$$

It follows using the slutsky's/Cramer's theorem that:

$$\sqrt{T}\hat{V}^{vec}_{\hat{\theta}_{1\lambda}} \Rightarrow \gamma vech\left[(\lambda Q)^{-1}\right]\mathcal{B}_1(\lambda)$$
$$\sqrt{T}\hat{V}^{vec}_{\hat{\theta}_{2\lambda}} \Rightarrow \gamma vech\left[((1-\lambda) Q)^{-1}\right]\left[\mathcal{B}_1(1) - \mathcal{B}_1(\lambda)\right]$$

Re-writing the above expressions gives us the following:

$$\sqrt{T}\hat{V}^{vec}_{\hat{\theta}_{1\lambda}} \Rightarrow \gamma vech\left[Q^{-1}\right]\frac{\mathcal{B}_1(\lambda)}{\lambda}$$
$$\sqrt{T}\hat{V}^{vec}_{\hat{\theta}_{2\lambda}} \Rightarrow \gamma vech\left[Q^{-1}\right]\frac{\mathcal{B}_1(1) - \mathcal{B}_1(\lambda)}{1-\lambda}.$$

Then,

$$\sqrt{T}\left(\hat{V}^{vec}_{\hat{\theta}_{1\lambda}} - \hat{V}^{vec}_{\hat{\theta}_{2\lambda}}\right) \Rightarrow \gamma vech\left[Q^{-1}\right]\frac{\mathcal{B}_1(\lambda) - \lambda\mathcal{B}_1(1)}{\lambda(1-\lambda)} \tag{6}$$

Additionally,

$$\sqrt{T}\left(\hat{V}^{vec}_{\hat{\theta}_{1\lambda}} - \hat{V}^{vec}_{\hat{\theta}_{2\lambda}}\right)' \Rightarrow \gamma\left(vech\left[Q^{-1}\right]\right)'\frac{\mathcal{B}_1(\lambda) - \lambda\mathcal{B}_1(1)}{\lambda(1-\lambda)} \tag{7}$$

Using a restricted variance estimator (i.e., under the null)[8]:

$$\widehat{V}_\lambda = \tfrac{1}{\lambda}\widehat{V}_{1\lambda} + \tfrac{1}{(1-\lambda)}\widehat{V}_{2\lambda} = \tfrac{1}{\lambda(1-\lambda)}\widehat{V} = \tfrac{1}{\lambda(1-\lambda)}\widehat{Var}\left(\sqrt{T}\widehat{v}_\epsilon vech\left[\left(T^{-1}\textstyle\sum_{t=1}^{T} X_t X_t'\right)^{-1}\right]\right)$$

Conditional on the regressors [9],

$$\widehat{V}_\lambda = \tfrac{1}{(\lambda)(1-\lambda)}vech\left[\left(T^{-1}\textstyle\sum_{t=1}^{T} X_t X_t'\right)^{-1}\right]\left(vech\left[\left(T^{-1}\textstyle\sum_{t=1}^{T} X_t X_t'\right)^{-1}\right]\right)'\widehat{Var}\left(\sqrt{T}\widehat{v}_\epsilon\right)$$

Using Assumption 4 and Lemma 1, we get:

$$\widehat{V}_\lambda = \tfrac{1}{(\lambda)(1-\lambda)}vech\left[\left(T^{-1}\textstyle\sum_{t=1}^{T} X_t X_t'\right)^{-1}\right]\left(vech\left[\left(T^{-1}\textstyle\sum_{t=1}^{T} X_t X_t'\right)^{-1}\right]\right)'\hat{\gamma}^2$$

$$\widehat{V}_\lambda \xrightarrow{p} \frac{1}{(\lambda)(1-\lambda)}vech\left[Q^{-1}\right]\left(vech[Q^{-1}]\right)'\gamma^2. \tag{8}$$

Combining (10), (11) and (12),

$$W_T \Rightarrow$$

$$\sup_{\lambda\in[\epsilon,1-\epsilon]} \frac{\left[\mathcal{B}_1(\lambda) - \lambda\mathcal{B}_1(1)\right]^2}{\lambda(1-\lambda)} \equiv \sup_{\lambda\in[\epsilon,1-\epsilon]} \mathcal{G}_1(\lambda)$$

□

Proof of Theorem 2:

We make use of the following result in our analysis:

Let A and X be symmetric $p \times p$ matrices, then

$$vech[AXA'] = H_p(A \otimes A)G_p vechX \tag{9}$$

Where H_p, known as the elimination matrix with the following dimensions, $p(p+1)/2 \times p^2$ and G_p, known as the duplication matrix with $p^2 \times p(p+1)/2$ - see Jan and Neudecker

[8] We could use a non-restricted variance estimator but we abstract from doing so to simplify the theoretical exposition and because using a split variance usually results in oversized tests.

[9] As mentioned earlier, the variance of any entity (be it the estimator itself or the variance of the estimator) is always conditional on knowing the distribution of regressors.

(1986) for details.
First, we have the following:

$$\sqrt{T}\hat{V}^{vec}_{\hat{\theta}_{i\lambda}} = \sqrt{T}vech\Big[\left(T^{-1}\textstyle\sum_{i\lambda} X_tX_t'\right)^{-1} \left(T^{-1}\textstyle\sum_{i\lambda} \widehat{\epsilon}_t^2 X_tX_t'\right) \left(T^{-1}\textstyle\sum_{i\lambda} X_tX_t'\right)^{-1} \Big], (i=1,2)$$

Applying the formula in (9), we get

$$\sqrt{T}\hat{V}^{vec}_{\hat{\theta}_{i\lambda}} = \underbrace{H_p\Big(\left(T^{-1}\textstyle\sum_{i\lambda} X_tX_t'\right)^{-1} \otimes \left(T^{-1}\textstyle\sum_{i\lambda} X_tX_t'\right)^{-1} \Big)G_p}_{I} \underbrace{vech\left(T^{-1/2}\textstyle\sum_{i\lambda} \widehat{\epsilon}_t^2 X_tX_t'\right)}_{II}, (i=1,2)$$

For i=1 and employing the Uniform Law of Large numbers, Assumption 2, Slutsky's theorem and the continuous mapping theorem, we get

$$I \xrightarrow{p} H_p\Big((\lambda Q)^{-1} \otimes (\lambda Q)^{-1} \Big)G_p$$

Similarly, for i=2, we have

$$I \xrightarrow{p} H_p\Big(\big((1-\lambda)Q\big)^{-1} \otimes ((1-\lambda)Q)^{-1} \Big)G_p$$

For i=1 and making use of the consistency of $T^{-1}\sum_{1\lambda} \hat{\epsilon}_t^2 X_tX_t'$ (Lemma 2), the following obtains:

$$T^{-1}\sum_{1\lambda} \hat{\epsilon}_t^2 X_tX_t' = \lambda v_{xe} + \mathcal{O}_p(1)$$

which can be re-written as,

$$T^{-1}\sum_{1\lambda} \big(\hat{\epsilon}_t^2 X_tX_t' - v_{xe}\big) = \mathcal{O}_p(1)$$

Then it follows that for i=1, $II = T^{-1/2}\sum_{1\lambda} vech\big(\hat{\epsilon}_t^2 X_tX_t' - v_{x\epsilon}\big)$

and for i=2, $II = T^{-1/2}\sum_{2\lambda} vech\big(\hat{\epsilon}_t^2 X_tX_t' - v_{x\epsilon}\big)$.

It follows from Assumption 3 and Lemma 1 that for i=1:

$$T^{-1/2}\sum_{1\lambda} vech\big(\hat{\epsilon}_t^2 X_tX_t' - v_{x\epsilon}\big) \Rightarrow \Omega^{1/2}\mathcal{B}_{p(p+1)/2}(\lambda)$$

In a similar fashion, for i=2,

$$T^{-1/2}\sum_{2\lambda} vech\big(\hat{\epsilon}_t^2 X_t X_t' - v_{x\epsilon}\big) \Rightarrow \Omega^{1/2}[\mathcal{B}_{p(p+1)/2}(1) - \mathcal{B}_{p(p+1)/2}(\lambda)]$$

It follows using slutsky's/Cramer's theorem that:

$$\sqrt{T}\hat{V}^{vec}_{\hat{\theta}_{1\lambda}} \Rightarrow \frac{1}{\lambda^2} H_p\Big(Q^{-1}\otimes Q^{-1}\Big) G_p \Omega^{1/2}\mathcal{B}_{p(p+1)/2}(\lambda)$$

$$\sqrt{T}\hat{V}^{vec}_{\hat{\theta}_{2\lambda}} \Rightarrow \frac{1}{(1-\lambda)^2} H_p\Big(Q^{-1}\otimes Q^{-1}\Big) G_p \Omega^{1/2}[\mathcal{B}_{p(p+1)/2}(1) - \mathcal{B}_{p(p+1)/2}(\lambda)]$$

$$\sqrt{T}(\hat{V}^{vec}_{\hat{\theta}_{2\lambda}} - \hat{V}^{vec}_{\hat{\theta}_{1\lambda}}) \Rightarrow$$
$$H_p\Big(Q^{-1}\otimes Q^{-1}\Big) G_p \left[\Omega^{1/2}\frac{(1-\lambda)^2\mathcal{B}_{p(p+1)/2}(\lambda) - \lambda^2[\mathcal{B}_{p(p+1)/2}(1) - \mathcal{B}_{p(p+1)/2}(\lambda)]}{\lambda^2(1-\lambda)^2}\right] \quad (10)$$

Additionally,

$$\sqrt{T}(\hat{V}^{vec}_{\hat{\theta}_{2\lambda}} - \hat{V}^{vec}_{\hat{\theta}_{1\lambda}})' \Rightarrow$$
$$\left[\Omega^{1/2}\frac{(1-\lambda)^2\mathcal{B}_{p(p+1)/2}(\lambda) - \lambda^2[\mathcal{B}_{p(p+1)/2}(1) - \mathcal{B}_{p(p+1)/2}(\lambda)]}{\lambda^2(1-\lambda)^2}\right]' G_p'\Big(Q^{-1}\otimes Q^{-1}\Big)' H_p' \quad (11)$$

Using a restricted variance estimator (i.e., under the null)[10]:

$$\widehat{V}_\lambda = \frac{1}{\lambda}\widehat{V}_{1\lambda} + \frac{1}{(1-\lambda)}\widehat{V}_{2\lambda} = \frac{1}{\lambda(1-\lambda)}\widehat{V} =$$

$$\frac{1}{\lambda(1-\lambda)}\widehat{Var}\left[H_p\left(T^{-1}\sum_{t=1}^T X_t X_t'\right)^{-1} \otimes \left(T^{-1}\sum_{t=1}^T X_t X_t'\right)^{-1} G_p vech\left(T^{-1/2}\sum_{t=1}^T \widehat{\epsilon}_t^2 X_t X_t'\right)\right]$$

Conditional on the regressors, the following matrix of dimensions, $p(p+1)/2 \times p(p+1)/2$, is constant.

$$\hat{P} \equiv H_p\left(T^{-1}\sum_{t=1}^T X_t X_t'\right)^{-1} \otimes \left(T^{-1}\sum_{t=1}^T X_t X_t'\right)^{-1} G_p$$

Then the variance expression simplifies to the following:

$$\frac{1}{\lambda(1-\lambda)}\hat{P}\cdot\widehat{Var}\left[vech(T^{-1/2}\sum_{t=1}^T \widehat{\epsilon}_t^2 X_t X_t')\right]\cdot\hat{P}' = \frac{1}{\lambda(1-\lambda)}\hat{P}\cdot\hat{\Omega}\cdot\hat{P}'$$

where $\hat{\Omega}$ is an estimate of Ω.

Then by slutsky's/Cramer's theorem and Assumptions 2 and 3,

$$\hat{P} \xrightarrow{p} P \equiv H_p\left(Q^{-1}\right)\otimes\left(Q^{-1}\right)G_p$$

[10]We could use a non-restricted variance estimator but we abstract from doing so to simplify the theoretical exposition and because using a split variance usually results in oversized tests.

$$\hat{V}_\lambda \xrightarrow{p} \frac{1}{\lambda(1-\lambda)} P \cdot \Omega \cdot P' \tag{12}$$

Define $J_{p(p+1)/2}(\lambda) \equiv (1-\lambda)^2 \mathcal{B}_{p(p+1)/2}(\lambda) - \lambda^2[\mathcal{B}_{p(p+1)/2}(1) - \mathcal{B}_{p(p+1)/2}(\lambda)]$

Combining (14), (15), and (16), we have

$$W_T \Rightarrow \sup_{\lambda \in [\epsilon, 1-\epsilon]} \frac{1}{\lambda^3(1-\lambda)^3} \left(J_{p(p+1)/2}(\lambda) \right)' \cdot \left(J_{p(p+1)/2}(\lambda) \right) \equiv \sup_{\lambda \in [\epsilon, 1-\epsilon]} \mathcal{G}_{(p)(p+1)/2}(\lambda)$$

□

References

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

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

Aue, Alexander, Siegfried Hormann, Lajos Horvath, and Matthew Reimherr. 2009. "Break Detection in the structure of multivariate time series models." *The Annals of Statistics*, 37: 4046–4087.

Bai, Jushan, and Pierre Perron. 1998. "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica*, 66(1): 47–78.

Corradi, Valentina. 1999. "Deciding between I(0) and I(1) via FLIL-based bounds." *Econometric Theory*, 15: 643–663.

Dette, Holger, and Dominik Wied. 2016. "Detecting relevant changes in time series models." *Journal of Royal Statistical Society Series B: Statistical Methodology.*

Eberlein, E. 1986. "On Strong Invariance Principles under dependent assumptions." *The Annals of Probability.*

Inclan, Carla, and George C. Tiao. 1994. "Use of Cummulative Sums of Squares for Reterospective Detection of Changes of Variance." *Journal of the American Statistical Association*, 89(427): 913–923.

Jan, Magnus, and H. Neudecker. 1986. "Symmetry, 0-1 Matrices and Jacobians." *Econometric Theory*, 157–190.

McLeish, D.L. 1975. "Invariance Principles for Dependent Variables." *Probability Theory and Related Fields*, 32: 165–178.

Perron, Pierre, Yohei Yamamoto, and Ying Zhou. 2020. "Testing jointly for structural changes in the error variance and coefficients of a linear regression model." *Quantitative Economics*, 11(3): 1019–1057.

Qu, Z., and P. Perron. 2007. "Estimating and Testing Structural Changes in Multivariate Regressions." *Econometrica*, 75: 459–502.