THE EFFECTS OF ADDITIVE OUTLIERS AND MEASUREMENT ERRORS WHEN TESTING FOR STRUCTURAL BREAKS IN VARIANCE

Paulo M. M. Rodrigues

Antonio Rubia

Please address correspondence to

Paulo M. M. Rodrigues

Economics and Research Department

Banco de Portugal, Av. Almirante Reis no. 71, 1150-012 Lisboa, Portugal;

Tel.: 351 21 313 0831, pmrodrigues@bportugal.pt

The Effects of Additive Outliers and Measurement Errors when Testing for Structural Breaks in Variance[*]

Paulo M.M. Rodrigues[†]

Banco de Portugal

Universidade Nova de Lisboa and CASEE

Antonio Rubia[‡]

Department of Financial Economics

University of Alicante

Abstract

This paper discusses the asymptotic and finite-sample properties of CUSUM-based tests for detecting structural breaks in volatility in the presence of stochastic contamination, such as additive outliers or measurement errors. This analysis is particularly relevant for financial data, on which these tests are commonly used to detect variance breaks. In particular, we focus on the tests by Inclán and Tiao [IT] (1994) and Kokoszka and Leipus [KL] (1998, 2000), which have been intensively used in the applied literature. Our results are extensible to related procedures. We show that the asymptotic distribution of the IT test can largely be affected by sample contamination, whereas the distribution of the KL test remains invariant. Furthermore, the break-point estimator of the KL test renders consistent estimates. In spite of the good large-sample properties of this test, large additive outliers tend to generate power distortions or wrong break-date estimates in small samples.

1   Introduction

There is much evidence that economic time-series are non-stationary when observed over long periods of time. Policy-regime shifts and exogenous factors may generate parameter instability in the underlying generating process, often leading to abrupt changes in time series dynamics. Recent literature in Financial Economics has addressed variance homogeneity, finding strong evidence of instability; see, *e.g.*, McConnell and Perez-Quirós (2000), Sensier and van Dijk (2004) and references therein. This topic is particularly important in financial markets because the second-order moment is central to Financial theory and its empirical applications.

The most popular statistical methods specifically designed to detect breaks in volatility are CUSUM-type tests. Into this category fall, among others, the tests by Pagan and Schwertz (1990), Inclán and Tiao (1994), Kokoszka and Leipus (2000), Chen, Choi and Zhou (2005), and Deng and Perron (2008a, 2008b). The ability of the CUSUM tests to identify structural changes depends critically on the empirical realism of the underlying assumptions. The most remarkable features of financial data include the existence of time-varying volatility patterns, and a tendency to generate abnormally large observations that cause similar effects as additive outliers.[1] Andreou and Ghysels (2002) illustrate the pervasive effect of persistent volatility on CUSUM tests experimentally. However, distortions that may arise from extreme observations, or other forms of stochastic contamination, have, to the best of our knowledge, not been addressed yet in the literature.

In this paper, we formally discuss the effects that sample contamination has on the asymptotic properties of CUSUM-type tests for detecting change points in variance and characterize the finite sample behavior by means of Monte Carlo simulations. Our theoretical discussion follows in a general framework in which additive outliers and/or measurement errors are treated as particular cases aiming to analyze the effects on i) the asymptotic distribution, ii) the consistency of the turning point estimator, and iii) the small-sample performance of CUSUM tests. Owing to their empirical relevance, special focus is placed on two well-known tests, but it should be stressed that the main conclusions are directly extensible to most testing procedures which are based on this framework. In particular, we study the parametric test suggested by Inclán and Tiao [IT] (1994), and the non-parametric generalization proposed by Kokoszka and Leipus [KL] (2000). These tests have been extensively applied on financial data; see, among others, Aggarwal, Inclán and Leal (1999), Andreou and Ghysels (2002, 2004), and Cuñado, Gómez-Biscarri and Pérez de Gracia (2006).

The results of our analysis can be summarized as follows. First, the IT test is not asymptotically invariant under stochastic contamination. It is biased towards serious overrejection – even in large samples – owing to the use of conservative critical values. In contrast, the non-parametric correction of the KL test ensures invariance and renders consistent estimates of the break-date. Our analysis reveals patterns which would be hard to explain in the absence of a formal theoretical analysis. For instance, the distribution of the IT test is more sensitive to a small likelihood of outliers than to a large probability of them, which is exactly the relevant case for empirical applications. Similarly, the asymptotic robustness of the KL test is not obvious *a priori*, since additive outliers may

---

[1] Outliers are discordant observations that seem to be far beyond the process that rules most observations. In financial markets, outliers are linked to rare shocks not related to the trading process, or abnormal flows of information arrivals. A well-known example is the market crash in October 19, 1987.

introduce asymptotic bias in least-squares based methods. In spite of the correct large-sample properties, using Monte Carlo analysis we observe that the KL test may suffer important power distortions in finite-samples when extreme outliers are present, which provide simple and straightforward reasons to explain the contradictory findings whenever the IT and KL tests are simultaneously applied (see, for instance, the analysis in Cuñado *et al.,* 2006, on data from emerging markets). Neglected outliers tend to bias the IT test towards finding a large number of breaks, whereas the KL test exhibits low power and tends to find few or no breaks at all.

The rest of the paper is organized as follows. Section 2 briefly outlines the test statistics which are analyzed in this paper. Section 3 derives the asymptotic properties under sample contamination, and provides a set of sufficient conditions to justify the results. Section 4 reports Monte Carlo results, which illustrate the small-sample behaviour of these tests and discusses the performance of the procedures in addressing variance homogeneity for US monthly stock returns. Section 5 summarizes and concludes. Finally, a technical appendix collects the proofs of the theoretical results presented in the paper.

2   Testing for structural breaks in variance

Assume that $\{r_t\}_{t=1}^{T}$ is the realization of a (zero-mean) stochastic process that verifies some restrictions which will be discussed later. The null hypothesis, $H_0 : Var(r_t) = \sigma^2$ with σ^2 constant over the entire sample, is tested against the alternative of single or multiple breaks of unknown location. The usual testing procedure infers the most likely break-position endogenously through cumulative sums of squared observations, r_t^2, as they provide an unbiased estimate of the unconditional variance. The key statistic is

$$D_T(k) = \left[\left(\sum_{t=1}^{k} r_t^2 \Big/ \sum_{t=1}^{T} r_t^2 \right) - k/T \right], \ \ k = 1, ..., T, \tag{1}$$

which can be viewed as an approximate likelihood ratio under some conditions. Under the alternative of a single break, the estimator of the break-date, say $\widehat{k}$, is determined as $argmax_k |D_T(k)|$. Whether this estimate is significant or not is then addressed through a test statistic based on $D_T(k)$ whose asymptotic distribution can be characterized as the *supremum* of a standard Brownian Bridge (SSBB henceforth) under suitable restrictions.

2.1   The Inclán and Tiao [IT] test

The IT test is a natural extension of CUSUM-type tests in regression models for the detection of shifts in variance, defined as

$$IT = \sqrt{T/2} \arg\max_{1 \le k \le T} |D_T(k)|. \tag{2}$$

Inclán and Tiao (1994) show that if, $r_t \sim iid\mathcal{N}(0, 1)$, then IT converges weakly to SSBB as $T \to \infty$. The IT test is initially intended to estimate the location of a single break as $\widehat{k} = \arg\max_{1 \le k \le T} |D_T(k)|$. However, a more general procedure based on the successive computation of (2) and the corresponding break-date estimation to gain power against the alternative hypothesis of multiple breaks can be considered. In particular, Inclán and Tiao

3

(1994) suggest the so-called Iterative Cumulative Sum of Squares (ICSS) method, which embeds the basic algorithm into an iterative scheme based on successive computations of (2) on different segments of the series, which are consecutively determined after a possible change point is detected.

2.2 The Kokoszka and Leipus [KL] test

The KL test statistic is defined as a suitably re-normalized version of $D_T(k)$, namely,

$$KL = T^{-1/2}\widehat{\mathcal{M}}_{4,T}^{-1/2}\arg\max_{1\leq k\leq T}|G_T(k)| \tag{3}$$

with $G_T(k) \equiv \sum_{t=1}^{k} r_t^2 - \left(\frac{k}{T}\right)\sum_{t=1}^{T} r_t^2$ and $\widehat{\mathcal{M}}_{4,T}$ being a consistent estimator of the long-run variance of $r_t^2 - E(r_t^2)$, i.e., the limit of $\frac{1}{T}E\left[\sum_{t=1}^{\infty}(r_t^2 - E(r_t^2))^2\right]$, say $\mathcal{M}_4 < \infty$. Since $G_T(k) = D_T(k)\sum_{t=1}^{T} r_t^2$, and noting that $\widehat{\sigma}^2 \equiv T^{-1}\sum_{t=1}^{T} r_t^2$, it follows that

$$KL = \sqrt{\frac{2\widehat{\sigma}^4}{\widehat{\mathcal{M}}_{4,T}}}IT. \tag{4}$$

The main purpose of the KL test is to weaken the Gaussian iid restriction of the IT test by using a model-free setting. Under fairly general conditions which do not hinge upon the particular distribution of the data, $\mathcal{M}_4$ can be estimated consistently using non-parametric techniques. Thus, the asymptotic distribution of the KL test is SSBB. The break-point estimator is defined as $\widehat{k} = \arg\max_{1\leq k\leq T}|G_T(k)|$, which is equivalent to $\arg\max_{1\leq k\leq T}|D_T(k)|$. Therefore, as in the IT test, the KL test can also be embedded into the same ICSS algorithm to gain power against multiple breaks.

3 Asymptotic theory

Consider first a data generating process [DGP] in which the main signal is perturbed with a stochastic contamination process that generates additive outliers and/or measurement errors, characterized under suitable restrictions. The objective is to define a process with similar statistical properties as those commonly found in financial and other economic time-series and to keep the assumptions to a minimum possible.

Assumption A1. *The observable data, $\{r_t, t \geq 1\}$, is generated from*

$$r_t = \mu + \varepsilon_t + \xi_t + \mathcal{B}_t[\lambda + \delta v_t] \tag{5}$$

where μ is some finite constant, ε_t is the regular component and ξ_t and $Z_t \equiv \mathcal{B}_t[\lambda + \delta v_t]$, are different sources of stochastic contamination.

Assumption A2. *The stochastic contamination components $\{\xi_t, Z_t\}$ observe the following properties:*
i) The measurement-error generator $\{\xi_t, t \geq 1\}$ is independent of ε_t and Z_t, and $\xi_t \sim$ iid $\left(0, \sigma_\xi^2\right)$ for some finite $\sigma_\xi \geq 0$. Also, $E\left(|\xi_t|^{4+\gamma}\right) < \infty$ for some $\gamma > 0$.
ii) The additive-outlier generator $\{Z_t, t \geq 1\}$, is independent of ε_t and ξ_t, with $\mathcal{B}_t$ being a discrete variable with support $(-1, 1, 0)$ and probabilities $\{p/2, p/2, 1 - p\}$. Furthermore,

$v_t \sim iid\,(0,1)$ *such that* $E\left(|v_t|^{4+\gamma}\right) < \infty$ *for some* $\gamma > 0$, $0 \le p < 1$, $0 < \lambda < \infty$, *and* $0 < \delta < \infty$.

Assumption A3. *The regular component* $\{\varepsilon_t, t \ge 1\}$ *observes the following properties:*
i) $E\left(\varepsilon_t\right) = 0$, $E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2 < \infty$.
ii) $sup_t E\left(|\varepsilon_t|^{4+\gamma}\right) < \infty$ *for some* $\gamma \ge 0$.
iii) ε_t *is strong mixing with mixing numbers* m_j *satisfying* $\sum_{j=1}^{\infty} m_j^{s/(s-2)} < \infty$ *for some* $s > 4$, *and* $\lim_{T\to\infty} E\left[T^{-1}\left(\sum [\varepsilon_t^2 - \sigma_\varepsilon^2]\right)^2\right] \equiv \mathcal{M}_4^\varepsilon$, *with* $0 < \mathcal{M}_4^\varepsilon < \infty$.

Assumption A3'. *Let* $\mathcal{F}_t$ *be the* σ-*field generated by* $\{\varepsilon_t, Z_t, \xi_t, \varepsilon_{t-1}, Z_{t-1}, ...\}$. *Then* $\{\varepsilon_t, \mathcal{F}_t\}_{t=1}^{\infty}$ *is a strictly stationary and ergodic martingale difference process and* $E\left(|\varepsilon_t|^{4+\gamma}\right) < \infty$ *for some* $\gamma \ge 0$.

Assumption A4. ε_t *is independent of* $\{\xi_t, Z_t\}$.

Some comments follow. Assumption A1 sets the basic DGP, in which measurement errors (ME) and/or additive outliers (AO) lead to the impossibility of observing the true signal ε_t.[2] In the financial literature, Z_t is usually referred to as a (discrete-time) stochastic jump process. In the econometric literature, a number of papers have focused on the effects of AOs through restricted forms of this general specification; see, *e.g.*, Franses and Haldrup (1994), van Dijk, Franses and Lucas (1999) and Vogelsang (1999). Assumption A2 assumes that AOs are generated independently of the regular component, which seems accurate for financial returns, as it captures extreme events which are unrelated to the normal trading process but which nevertheless are able to influence the observable series. Similarly, the ME component is assumed to be exogenous, which does not seem a particularly restrictive condition in practice. Since ξ_t and Z_t are bounded in probability, ε_t is not perturbed by arbitrarily large values.[3] Condition A3 is fairly general and standard in the literature. It allows for finite-order ARMA structures and/or several time-varying volatility processes, such as stationary GARCH-type and stochastic volatility models. Condition A3' may be sufficient when the series are uncorrelated but not independent, as is mostly the case in financial time-series. Finally, A4 is a maintained assumption in related studies, such as Franses and Haldrup (1994) and van Dijk *et al.* (1999). Its practical purpose is to allow us to analyze the effects of stochastic contamination in a model-free framework. We shall comment on the effects of weakening this assumption later on.

3.1   Asymptotic distribution of the test statistics

In this section, the asymptotic distributions of the IT and KL test statistics under additive outliers and/or measurement errors are formally derived. We denote '$\Rightarrow$' as the weak convergence of probability measures in $D\,[0,1]$, '$\xrightarrow{p}$' as convergence in probability, $W\left(\tau\right)$ represents a standard Wiener process on $\tau \in [0,1]$, $W^*\left(\tau\right) = W\left(\tau\right) - rW\left(1\right)$ is a standard Brownian Bridge, $[\cdot]$ is the integer function, and $tr\left(\cdot\right)$ denotes the trace of a matrix.

---

[2]It is widely accepted that AOs are part of the return's generating process. Non-synchronous and thin-trading may generate measurement errors, particularly, in data recorded from emerging markets.

[3]Alternatively, a more extreme (but somewhat less general) form of contamination in which outliers are allowed to be arbitrarily large could be considered as well. The analysis of the performance of the CUSUM tests, and the formalization of robustified alternatives, constitute interesting topics for future research. We thank an anonymous referee for bringing this issue to our attention.

Lemma 3.1. *Define the random vector $\Pi_t = \left(\varepsilon_t^2, Z_t^2, \xi_t^2, 2\varepsilon_t Z_t, 2\varepsilon_t \xi_t, 2\xi_t Z_t\right)'$ and assume A1-A4 hold true. Then, as $T \to \infty$, $\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \left(\Pi_t - E\left(\Pi_t\right)\right) \Rightarrow \Omega^{1/2} \mathbf{W}\left(\tau\right)$ in $D\left[0,1\right]^6$ and uniformly in $\tau \in \left[0,1\right]$, where $\mathbf{W}\left(\tau\right)$ is a multivariate standard Wiener process with covariance matrix $\Omega = \left\{\omega_{ii}\right\}$, $\omega_{11} = \mathcal{M}_4^\varepsilon$; $\omega_{22} = p\left[\lambda^4 - \lambda^2 + \delta^4 \mu_4^v + 2\lambda\delta^2\left(3\lambda + \delta\mu_3^v\right) - \delta^2\right]$; $\omega_{33} = \mu_4^\xi - \sigma_\xi^2$; $\omega_{44} = 4\sigma_\varepsilon^2 p\left(\lambda^2 + \delta^2\right)$; $\omega_{55} = 4\sigma_\varepsilon^2 \sigma_\xi^2$; $\omega_{66} = 4\sigma_\varepsilon^2 \sigma_\xi^2$, and $\omega_{ij} = 0$ for $i \neq j$; where $E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2$, $\mu_j^v = E\left(v_t^j\right)$, $\mu_j^\xi = E\left(v_t^j\right)$.*

Lemma 3.2. *Denote $\tilde{r}_t = r_t - \widehat{\mu}_T$, where $\widehat{\mu}_T$ is a $\sqrt{T}$-consistent estimator of μ. Under the conditions of Lemma 3.1, as $T \to \infty$, it can be established that* (i) $\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \left[\tilde{r}_t^2 - E\left(r_t^2\right)\right] \Rightarrow \mathcal{M}_4^{1/2} W\left(\tau\right)$; (ii) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\tilde{r}_t^2 - E\left(r_t^2\right)\right] \Rightarrow \mathcal{M}_4^{1/2} W\left(1\right)$ *and* (iii) $\frac{1}{T} \sum_{t=1}^{T} \tilde{r}_t^2 \xrightarrow{p} \sigma_\varepsilon^2 + p\left(\lambda^2 + \delta^2\right) + \sigma_\xi^2$, *for any $r \in \left[0,1\right]$, where $\mathcal{M}_4 \equiv Var\left(r_t^2 - E\left(r_t^2\right)\right) = tr\left(\Omega\right)$.*

Lemmas 3.1 and 3.2 state the convergence of the functionals involved in the IT and KL tests and clarify the dependence on the main driving parameter (see the technical appendix for details). The following theorem presents the asymptotic distribution of the tests.

Theorem 3.1. *Let the IT and KL test statistics be defined as in (2) and (3), respectively, and let $\widehat{\mathcal{M}}_{4,T}$ be a consistent estimator of the long-run variance $\mathcal{M}_4$. Then, under the conditions of Lemma 3.1 and as $T \to \infty$,*

$$
\begin{aligned}
IT &\Rightarrow \frac{\mathcal{M}_4^{1/2}}{\sqrt{2}\left(\sigma_\varepsilon^2 + p\lambda^2 + p\delta^2 + \sigma_\xi^2\right)} \sup_{\tau \in [0,1]} \left|W^*\left(\tau\right)\right| \\
KL &\Rightarrow \sup_{\tau \in [0,1]} \left|W^*\left(\tau\right)\right|.
\end{aligned}
\tag{6}
$$

Corollary 3.1. *If only additive outliers contaminate the sample, then Theorem 3.1 trivially holds with $\sigma_\xi = \mu_4^\xi = 0$, whereas if only measurement errors are present, i.e., $p = 0$, then Theorem 3.1 trivially holds by setting all parameters related to the AOs equal to zero.*

The proofs of Theorem 3.1 and its corollary follow directly from Lemmas 3.1 and 3.2 and the continuous mapping theorem (see the appendix). Theorem 3.1 states formally one of the theoretical results of this paper, and has important implications for empirical purposes. The asymptotic distribution of the IT test is not invariant and it turns out to be heavily influenced by the characteristics of the contamination process, since this introduces non-Gaussian features, such as excess kurtosis. The critical values from the correct asymptotic distribution will be larger whenever $tr\left(\Omega\right) > 2E\left(r_t^2\right)^2$, and smaller otherwise. Since, the characteristics of the contaminating process are not observable, the IT test becomes infeasible under outlying observations and/or measurement errors. In sharp contrast, the KL test is scaled with a non-parametric HAC-type estimator of the long-run variance that succeeds in making the procedure robust to contamination, ensuring its convergence to the SSBB. Vogelsang (1999) finds similar results in the different context of unit root testing, suggesting the use of the Phillips-Perron test procedure (which builds on the autoregressive spectral density estimation of the long-run variance parameter) to deal with AOs.

Remark 3.1: The diagonality of the asymptotic covariance matrix Ω which characterizes $\mathcal{M}_4$ follows directly from Assumption A4. We may allow for dependence between some

measurable function of ε_t, say $\mathcal{V}_1(\varepsilon_t)$, and lagged values of a function $\mathcal{V}_2(Z_t, \xi_t)$, provided conditions A3 or A3' are still fulfilled. Relevant examples of $\mathcal{V}_1$ and $\mathcal{V}_2$ in this context are quadratic and absolute-value functions, given that the volatility process may be affected by lagged outliers. Define $\bar{\bar{\Pi}}_t = \Pi_t - E(\Pi_t)$ and let $\Omega^* = \lim_{T\to\infty} \frac{1}{T} E\left(\left(\sum_{t=1}^T \bar{\bar{\Pi}}_t \right) \left(\sum_{t=1}^T \bar{\bar{\Pi}}_t' \right) \right)$. If A4 is replaced by the assumption that Ω^* is a finite positive definite matrix such that $\Omega^* = \Lambda\Lambda'$, then it follows as $T \to \infty$ that $\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \bar{\bar{\Pi}}_t \Rightarrow \Lambda \mathbf{W}(\tau)$, with this result generalizing Lemma 3.1 in an obvious way. The diagonal elements of Ω^* are those of Ω, whereas the (non-zero) off-diagonal elements of this matrix depend specifically on the covariance structure related to $\mathcal{V}_1$ and $\mathcal{V}_2$, and hence are model-dependent. It also follows that $\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} [\tilde{r}_t^2 - E(r_t^2)] \Rightarrow \mathbf{1}'\Lambda \mathbf{W}(\tau)$, and $\frac{1}{T} \sum_{t=1}^T \tilde{r}_t^2 \xrightarrow{p} \mathbf{1}'E(\Pi_t\Pi_t')\mathbf{1}$, with $\mathbf{1}$ being a conformable vector of ones. As in Theorem 3.1, it can be shown that $IT \Rightarrow \zeta_{\mathcal{V}_1,\mathcal{V}_2} \sup_{\tau\in[0,1]} |W^*(\tau)|$, with $\zeta_{\mathcal{V}_1,\mathcal{V}_2} = \sqrt{\mathbf{1}'\Omega^*\mathbf{1}} \left(\sqrt{2} \left(\mathbf{1}'E(\Pi_t\Pi_t')\mathbf{1} \right) \right)^{-1}$, and again $KL \Rightarrow \sup_{\tau\in[0,1]} |W^*(\tau)|$. This issue will be analyzed more carefully in the Monte Carlo section.

3.2   Consistency of the change-point estimator

We now discuss the ability of the tests to consistently estimate the location of an unknown turning point under the alternative hypothesis. We initially assume that only a single break affects the dynamics of the regular component ε_t, as there is no practical sense in considering breaks in the contaminating structure. It is also necessary to make additional assumptions to characterize the nature of the structural break and introduce further notation. Denote κ^*, $0 < \kappa^* < 1$, as the break fraction, such that a shift occurs at time $k^* + 1$, $k^* = [T\kappa^*]$, and let $\varepsilon_{1t} = \{\varepsilon_t\}_{t=1}^{k^*}$ and $\varepsilon_{2t} = \{\varepsilon_t\}_{t=k^*+1}^T$ be the pre- and post-break sub-samples such that $E(\varepsilon_{1t}^2) = \sigma_{1\varepsilon}^2$ and $E(\varepsilon_{2t}^2) = \sigma_{1\varepsilon}^2 + \Delta > 0$. The break is of (finite) magnitude Δ and may be originated in the conditional or unconditional variance of the regular series. In both cases, the shift in the variance of ε_t leads to a shift of the same magnitude in the variance of r_t, i.e., $\Delta = \text{Var}(r_{t,t\le[T\kappa^*]}) - \text{Var}(r_{t,t>[T\kappa^*]})$. This property allows testing for changes in the unobservable component ε_t using the observed series r_t instead, but introduces inefficiency. Note that both the IT and KL test procedures yield the same break-fraction estimation, namely $\hat{\kappa} = T^{-1} \arg\max_{1\le k\le T} |D_T(k)|$. Our interest is to analyze if $\hat{\kappa} \xrightarrow{p} \kappa^*$ under the set of assumptions considered and the following additional condition.

Assumption A5. The sequence $\{\varepsilon_t\}_{t=1}^\infty$ verifies $i)$ $\sup_t E(\varepsilon_t^4) < \infty$ for all t, and $ii)$ $\text{Cov}(\varepsilon_t^2, \varepsilon_j^2) = O(\rho^{|t-j|})$ for all $1 \le t, j \le T$ and some $0 \le \rho < 1$.

Condition A5 is embedded in A3 or A3' when there are no breaks. With $i)$ we rule out shifts which dramatically change the statistical properties of the process as considered under the null, such as parameter instability leading to diverging moments up to the fourth order. Condition $ii)$ restricts the covariance structure of the time-series. Although ε_t is not stationary under parameter instability, we still require that the covariance between distant observations decays towards zero at a suitable rate. Note that $ii)$ holds trivially for independent series, as well as for short-memory series. More importantly, $ii)$ may be weakened considerably, as consistency can be proven under the more general condition $\lim_{T\to\infty} T^{-2} \sum_{t=1}^T \sum_j^t Cov(\varepsilon_t^2, \varepsilon_j^2) = 0$, which may follow under suitable mixing conditions

and allows for different rates of decay in the covariances. Convergence in probability for the estimator $\widehat{\kappa}$ is provided as a theorem below.

Theorem 3.2. *Consider $\{r_t\}_{t=1}^T$ as defined in A1 such that A2 and A4 hold true. Assume that the unconditional variance of the regular component shifts from σ_ε^2 to $\sigma_\varepsilon^2 + \Delta > 0$, $0 < |\Delta| < \infty$, at some time $k^* = [T\kappa^*]$ for some $\kappa^* \in (0,1)$ such that A5 holds true. Let $\widehat{\kappa} = T^{-1} \arg\max_{1 \leq k \leq T} |D_T(k)|$. Then, for an arbitrary $\epsilon > 0$ and some constant C it follows that*

$$\Pr\left(|\widehat{\kappa} - \kappa^*| > \epsilon\right) \leq \frac{C}{\epsilon^2 \Delta^2 \sqrt{T}} \tag{7}$$

and, therefore, $\widehat{\kappa} \xrightarrow{p} \kappa^*$ as the sample size is allowed to diverge.

Remark 3.2: Consistency holds for any estimator based on $\arg\max_{1 \leq k \leq T} |D_T(k)|$. If we allow for cross-dependences, as in Remark 3.1, then $\widehat{\kappa} \xrightarrow{p} \kappa^*$ holds generally if $\sup_t E\left(\varepsilon_t^4\right) < \infty$ and $\lim_{T \to \infty} T^{-2} \sum_{t=1}^T \sum_j^t Cov\left(r_t^2, r_j^2\right) = 0$.

The proof of consistency uses the Hájek-Rénji inequality in Kokoszka and Leipus, (2000), see the appendix for details. For a fixed shift Δ, the bias $|\widehat{\kappa} - \kappa^*|$ can be shown to be $O_p\left(T^{-1}\right)$ and, therefore, $\widehat{\kappa}$ is super-consistent, a standard result in this literature. Consistency guarantees that the KL test not only uses the correct critical values, but also that these are applied on the correct location when the sample grows unbounded. Furthermore, Theorem 3.2 provides insight on how the representative characteristics of the DGP affect the bias. In particular, the degree of serial dependence, the time of the break, and the excess of variability generated by the contaminating process increase the size of C and will make it more difficult to locate the break-position correctly in finite samples (see appendix for further details). On the other hand, the sample size and the magnitude of the shift help to reduce the bias. This statistical tension completely disappears when $T \to \infty$, but (7) indicates the sources of small sample distortions for finite T. This issue will be studied in greater detail in the Monte Carlo section below.

The extension of the single-break analysis towards the detection of multiple changes in variance is usually done through the ICSS algorithm suggested by IT; see Chen *et al.* (2005). The procedure starts by applying the single break-point detection over the entire sample. If a break is detected, the whole sample is divided into two subsamples, and the testing procedure is applied again in each subsample. The process is repeated until no changes are detected, yielding an estimate of the number of breaks, say $\widehat{m}$. Since, from Theorems 3.1 and 3.2 the KL test consistently estimates and identifies a single break-point, the ICSS algorithm based on this test will also consistently estimate the unknown number of breaks, say m, in the multiple break context. We provide this result as a proposition below (the proof follows along the lines in Bai, 1998, and is therefore omitted to save space).

Proposition 3.1: *Suppose that the conditions in Theorem 3.2 hold true and that there is a number $m < T$ of shifts in the unconditional variance of the regular process. In particular, assume that for any $j = 1, ..., m$, the variance shifts from σ_ε^2 to $\sigma_\varepsilon^2 + \Delta_j > 0$, $|\Delta_j| < \infty$, at time $k_j^* = \left[T\kappa_j^*\right]$, $\kappa_j^* \in (0,1)$, such that $k_j^* - k_{j-1}^*$ ($k_0^* \equiv 0$) includes a non-trivial set of observations, and A5 holds true. Let $\widehat{m}$ be the number of breaks inferred from a sequential procedure such as the ICSS approach, then $\widehat{m} \xrightarrow{p} m$.*

4 Finite-sample analysis

4.1 Monte Carlo analysis

Given its relevance, and in order to save space, in this analysis we center our attention on additive outlier effects and focus only on the single-break case. Several experiments are considered to study possible small-sample size departures related to outliers when the regular component is A) an iid process, and B) exhibits time-varying volatility patterns. In addition, C) we also address the consistency of turning point estimation in small-samples.

A) Empirical size: additive outliers and independent observations.

We first assume that the regular component ε_t is iid to isolate the effects of outliers. We generate simulated paths for $r_t = \varepsilon_t + \mathcal{B}_t\left[\lambda + \delta v_t\right]$, $\varepsilon_t = \sigma \eta_t$, with η_t and $v_t \sim iid\mathcal{N}(0,1)$. The discrete variable $\mathcal{B}_t = (-1, 1, 0)$ with the grid of probabilities $p = \{0, 0.01, ..., 0.50\}$, and increments of 0.01 is used. We set $\lambda \in \{0.5, 1, ..., 5\}$, and increments of 0.5, covering a wide set of values including many relevant ones for empirical purposes. We initially set $\delta = \sigma = 1$. The sample size is $T = 1000$ and we repeat the simulation process 25000 times for any combination of the analyzed values. The IT and the KL tests are computed using the simulated series, and the corresponding statistics compared to the 95% percentile from the SSBB (*i.e.*, 1.36). The rejection rates of the null hypothesis are depicted in Figures 1 and 2.

[Insert Figures 1, 2 about here]

As discussed in the theoretical section, the distribution of the IT test is strongly affected by the parameters that characterize the dynamics of the outlier process. We observe that extreme values lead to large size departures, specially for large, infrequent values (small p and large λ), which is precisely the type of process that is to be expected in real financial data. These values generate large excess kurtosis and lead the IT test to over-reject.[4] Although not reported here, in order to save space, the distortion in size amplifies as δ increases (given that kurtosis depends on this parameter as well). In contrast, the KL test shows a flat, uniform distribution for the empirical rate of rejection which does not depend on nuisance parameters.

B) Empirical size: additive outliers and time-varying volatility.

We now turn our attention to the case in which the regular component exhibits time-varying volatility patterns. Owing to its empirical relevance, we assume that ε_t follows a GARCH(1,1) model, such as

$$
\begin{aligned}
r_t &= \sigma_t \eta_t + Z_t; \quad \eta_t \sim iid\mathcal{N}(0,1) \\
\sigma_t^2 &= \omega + \alpha \psi_{t-1}^2 + \beta \sigma_{t-1}^2,
\end{aligned}
\tag{8}
$$

as a function of the variable ψ_t. In the context of outliers, there are two possible specifications, depending on whether they affect the level of the series (*level outliers*), or affect both the level and the variance (*volatility outliers*). In the first case, σ_t^2 is independent of

---

[4]Relatively large values of p lead to undersized tests. This may seem surprising, but from the analysis of Theorem 3.1 we note that large values of p lead to $tr\left(\Omega\right) < 2E\left(r_t^2\right)^2$, hence undersizing the test. Such a degree of heterogeneity, however, is unlikely observed in practice.

Z_t, and therefore we set $\psi_t = \sigma_t \eta_t \equiv \varepsilon_t$. Thus, observed returns result from the convolution of a jump-process and a standard GARCH model. This is the case explicitly studied in the theoretical section, and the DGP considered in most empirical applications. Alternatively, if σ_t^2 is also perturbed by AOs, then $\psi_t = r_t$ and the GARCH model includes jumps, a far more complex type of non-linear volatility process. We shall analyze both possibilities in our simulations.

First, consider the level-outlier case with AOs affecting only the conditional mean. We normalize the unconditional variance to unity by setting $\omega = (1 - \alpha - \beta)$, and set different values for (α, β). In particular, we consider the same DGPs as in Andreou and Ghysels (2002) - a low persistent $(\alpha = 0.10, \beta = 0.50)$ [GARCH1] and a high-persistent GARCH model $(\alpha = 0.10, \beta = 0.80)$ [GARCH2] - to make comparisons with their results. Simulations are performed as in experiment A), with the long-run variance being computed, for the KL test using a Newey-West estimator with Bartlett kernel and a deterministic bandwidth selection procedure. The empirical rates of rejection are shown in Figures 3, 4 and 5.

<div align="center">[Insert Figure 3, 4, 5 about here]</div>

Figure 3 confirms the theoretical results for the IT test presented in the previous section.[5] Since GARCH-type dependence originates excess kurtosis, time-varying volatility suffices to bias the IT test even if $p = 0$. When $p > 0$, rare extreme events (low p and large λ) considerably increase total kurtosis, leading to even larger size distortions.

Figures 4 and 5 show the empirical size of the KL test given GARCH1 and GARCH2 errors, respectively. The most striking feature is related to the effect generated by different degrees of persistence in volatility, and not by the presence of outliers. Remarkably, in the absence of outliers, the KL test suffers from important small-sample size distortions for large values of $\alpha + \beta$. This feature was already reported in the simulations in Andreou and Ghysels (2002) and, therefore, cannot be attributed solely to outliers. This is actually a finite-sample distortion related to the fact that the HAC estimator of the long-run variance suffers large small-sample bias in strongly persistent data. Further simulations (not reported here) show that this bias worsens for smaller samples. The existence of AOs does not worsen the behavior of the KL test with respect to the case $p = 0$ and, therefore, the major distortions observed are solely attributable to dependence patterns in volatility.

When outliers affect both the conditional mean and variance, *i.e.*, setting $\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$ in (8), the departures from the nominal size of the IT test are even larger than before. For instance, in the case of the GARCH2 model we do not observe empirical sizes inferior to 50% (results are available upon request). Excess kurtosis in r_t is now even greater as a result of the positive correlation between ε_t^2 and the lagged values of Z_t^2. In the case of the KL test, the existence of outliers does not have negative effects on the empirical size.

C) Finite-sample properties: consistency of the break-point estimator

As in related studies (see Chen *et al.* 2005), the last experiment aims to evaluate the average size of the estimation bias $(\widehat{\kappa} - \kappa^*)$, and the corresponding standard error of the break-point estimate (*i.e.*, efficiency) in finite samples. We set the turning-point fractions

---

[5] For the IT test, the results for the GARCH1 model are qualitatively similar to those from GARCH2. We therefore do not present them, but they are available upon request.

$\kappa^* = \{0.25, 0.50, 075\}$ and normalize to unity the pre-break unconditional variance para-meter, $\sigma_1 = 1$. In addition, we consider several values for Δ, and perform simulations with a DGP in which the regular component is iid, or follows GARCH1 or GARCH2 errors. For the sake of conciseness, we summarize the results of this experiment for several values of (λ, p), $\delta = 1$, a relatively large shift $\Delta = 0.50$, and for the three forms of dependence in ε_t in Table 1, showing the average value of $\widehat{\kappa}$ and its standard error in a sample of $T = 1000$.

[**Insert Table 1 about here**]

The CUSUM principle consistently detects structural breaks, since increasing T and/or the magnitude of the shift reduces the estimation bias and the standard error of $\widehat{\kappa}$. How-ever, as expected the properties of $\widehat{\kappa}$ prove sensitive to the characteristics of outliers in finite samples, particularly, the size of λ. Large values of this parameter bias $\widehat{\kappa}$ towards $1/2$ (as in the no break case), and considerably increase the standard error of the es-timates. Extreme additive outliers may generate large biases in finite samples, even if they occur with small probability. The reason is that the additional variability generated by multiple outliers can mask the true position of the variance shift, and bias the least-squares estimates. As a result, although the IT test will tend to find spurious breaks from using over-conservative critical values, the KL test may be biased towards non-detection because the correct critical values may be applied on wrongly estimated turning-points, thereby rejecting the null.

We analyze this effect through further experimentation. We only discuss the main results without presenting tables in order to save space. Consider the most favorable case for the KL test in which ε_t is iid, and assume that a large single shift increases the uncondi-tional variance from 1 to 1.5 in a sample of 1000 observations, with $\kappa^* = \{0.25, 0.50, 0.75\}$. In the absence of outliers, the average value of $\widehat{\kappa}$ is always in the neighborhood of the true κ^*, and the probability of rejecting the false null is nearly 100% for a 5% nominal size. In sharp contrast, if the series is randomly contaminated with large, infrequent AOs (*e.g.*, setting $\lambda = 5$ and $p = 0.10$) the average value of $\widehat{\kappa}$ is biased towards the middle of the sample, $E(\widehat{\kappa}) = \{0.41, 0.51, 0.41\}$ and, even more importantly, the probability of rejecting the false null dramatically collapses to $\{21.1\%, 41.2\%, 25.7\%\}$. Even though these biases are entirely attributable to finite-sample effects, and will eventually vanish as the sample size diverges, the experiment highlights the direction and extent of the small sample bias of the KL test.

4.2 Empirical application: variance stability in the US market

It is interesting to analyze empirical data in the light of our theoretical and experimental analysis. In particular, we apply the IT and KL statistics to test for variance homogeneity on monthly returns from the US market from March 1885 to December 2001 available from William G. Schwert's website. We have a large number of observations (1398) spanning a period in which it is widely acknowledged that the US stock market experienced, at least, a period of abnormally high volatility during the Great Depression (GD), which the tests should be able to detect. Note that serial dependence in volatility is considerably weakened at the monthly frequency, so we do not expect large distortions in the size of the tests due to persistence in conditional variance. Furthermore, monthly returns time-series exhibit a large degree of leptokurtosis and non-normality owing to extreme values, which are expected to generate distortions according to our previous analysis. In addition to

October 1987, the most influential observations are related to episodes of international crisis, World War I and II, and the energy crisis of 1973. Overall, the sample provides us with a perfect ground to analyze the empirical performance of the tests.

We compute the IT and KL tests on squared- and absolute-valued series. Although both transformations track the dynamics of the second-order moment, the information conveyed by those transformations is not necessarily the same, and as a result, the inferred number of breaks, and even their estimated location, may significantly vary from using one proxy or the other; see Andreou and Ghysels (2002) and Cuñado *et al.* (2006). Since influential outliers are expected to have a detrimental impact on the ability of the tests, we apply a trimming procedure to exclude the largest observations and check the robustness of any preliminary conclusion. Many papers in applied finance control for outliers by filtering top percentiles, observations that lie beyond some pre-determined level, or simply by removing specific observations, such as October 1987. Along with the non-trimming fraction (0%), we also apply conservative trimming fractions ranging from 0.5% (thus filtering only 7 observations) to 2.5% (35 observations) to remove most extreme observations. If conclusions change dramatically after removing a few observations, then preliminary conclusions may be spurious and purely driven by outliers.

[Insert Table 2 about here]

The estimated break-locations are reported in Table 2. As expected from our previous analysis, the IT test tends to identify a relatively large number of breaks when applied on the original series, whereas the KL test tends to find significantly smaller numbers. Similar evidence is observed, for instance, in Cuñado *et al.* (2006). More specifically, the IT test on absolute-valued (*squared*) returns finds up to 12(9) breaks at a 5% nominal level, whereas the KL test cannot reject the null hypothesis in any of those series.[6] Likely owing to the pervasive effects of outliers, the IT test seems to identify short-lived structural breaks (*e.g.*, 1940.04-1940.06), and outliers (*e.g.*, the 1987 crash) as structural breaks. The KL test is unable to detect breaks around the GD.

The spuriousness of these results is evident after controlling for outliers, since the main conclusions dramatically change after filtering just a small set of observations. It suffices to remove the most influential 0.5% to dramatically reduce the number of breaks found by the IT test (the IT test on $|r_t|$ only finds breaks around the beginning and end of the GD and the end of the 19th century) and to allow the KL test on $|r_t|$ to detect the GD instability. This empirical exercise perfectly illustrates the main conclusions discussed theoretically and analytically in the previous sections.

5   Conclusions

In this paper, the size properties of CUSUM-type tests for detecting structural breaks in variance when the series of interest include some of the most relevant features that characterize financial data were analysed. Our special focus has been on additive outliers, which prove able to generate large size distortions in these tests. The most sensitive procedure is

---

[6]When applying the tests on $|r_t|$ and r_t^2 with no previous filtering, the first potential break dates found are 1940:05 and 1940:06, respectively. In the case of the IT test, both dates are identified as break-points, so that the ICSS algorithm continues. The KL test, on the other hand, fails to reject the null, and therefore the iterative procedure stops. These results remain valid even when using a 10% nominal level.

the one by Inclán and Tiao (1994), which should be applied with considerable caution on financial data. On the other hand, the procedure by Kokoszka and Leipus (2000) exhibits much better behaviour, at least under the conditions considered in this paper. In particular, the asymptotic distribution of the test statistic is the standard one, and the estimator of the break point is consistent. However, in finite-samples large size distortions due to the presence of outliers may be observed. As in the case of strongly-persistent volatility patterns, certain characteristics of the empirical data generating process in financial time series may cause major distortions in the small-sample performance of CUSUM-type tests.

Consequently, a question of empirical relevance is what to do with extreme anomalous observations. Bai (1998) proposed the use of robust procedures in the context of regressions with structural changes and additive outliers. In this spirit Fiteni (2004) has recently proposed the use of bounded-influence estimators. These methods may outperform least-squares based estimators under possible contaminated distributions, as they are specifically designed to be used even under arbitrarily large outliers. Since the standard CUSUM test for detecting structural breaks in variance build on least-squares (or maximum likelihood) estimates , the use of bounded-influence estimators may provide further improvements of the small-sample performance of this approach. The findings and general discussion in our paper support the empirical pursuit of this interesting question in future research.

A   Appendix

Proof of Lemma 3.1.
We consider $\mu = 0$ in A1 for simplicity but without loss of generality. Let $I_{T,\tau} = \sum_{t=1}^{[T\tau]} [\Pi_t - E(\Pi_t)]$ and denote the j-th entry of this vector as $I_{j,\tau}$. Under A2 and A4, the covariance matrix of $I_{T,\tau}$, Ω, is diagonal because the $I_{j,\tau}$ terms are uncorrelated, and $I_{1,\tau}, I_{4,\tau}$ and $I_{5,\tau}$ satisfy a functional central limit theorem (FCLT) for mixing sequences (martingale differences) under A3 (A3'), while $I_{4,\tau}, I_{5,\tau}$ and $I_{6,\tau}$ verify directly the FCLT from Donsker's lemma under A2, c.f. White (2000) and Deng and Perron (2008a). Hence, $\frac{1}{\sqrt{T}} I_{j,\tau} \Rightarrow \sqrt{\omega_{jj}} W(\tau)$, where ω_{jj} is the j-th element of the main diagonal in Ω. It follows from the Gaussian properties of the Wiener process that $\frac{1}{\sqrt{T}} I_{T,\tau} \Rightarrow \Omega^{1/2} \mathbf{W}(r)$, where $\mathbf{W}(r)$ is a 6-dimensional Wiener process. Since $r_t^2 = \mathbf{1}'\Pi_t$, with $\mathbf{1}$ a vector of ones in $\mathbb{R}^6$, and $E(r_t^2) = \mathbf{1}'E(\Pi_t)$, then $\sum_{t=1}^{[T\tau]} r_t^2 - E(r_t^2) = \sum_{j=1}^6 I_{j,\tau}$. The Cramer-Rao device completes the proof. ∎

Proof of Lemma 3.2.
Since $\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} [\tilde{r}_t^2 - E(r_t^2)] = \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \mathbf{1}'[\Pi_t - E(\Pi_t)] + o_p(1)$, it follows that the limit distribution of the functional converges weakly to the distribution of $\mathbf{1}'\mathbf{W}(\tau) = W(\tau)$, a standard Wiener process, with scalar variance $\mathbf{1}'\Omega\mathbf{1} = tr(\Omega)$. This yields the required result. Part (ii) of the lemma is immediate for $\tau = 1$, and part (iii) follows, similar to Lemma 3.1, from applying the weak law of large numbers; see White (2000). ∎

Proof of Theorem 3.1.

Observe that $D_T(k)$ can be rewritten as $C_T(T)^{-1}[C_T(k) - (k/T)C_T(T)] = C_T(T)^{-1}G_T(k) + o_p(1)$ where $C_T(k) = \sum_{t=1}^{k} r_t^2$. Therefore, $\frac{1}{\sqrt{T}}D_T(k) = \left[\frac{C_T(T)}{T}\right]^{-1}\left[\frac{1}{\sqrt{T}}G_T(k)\right] + o_p(1)$. From i) and ii) of Lemma 3.2 it follows that $\frac{1}{\sqrt{T}}\left[C_T(k) - \left(\frac{k}{T}\right)C_T(T)\right] \Rightarrow \sqrt{\mathcal{M}_4}W^*(\tau)$ where from Lemma 3.2iii), $\frac{C_T(T)}{T} \xrightarrow{p} Var(r_t) = \sigma_\varepsilon^2 + p(\lambda^2 + \delta^2) + \sigma_\xi^2$. From this result and the continuos mapping theorem, it follows that

$$\arg\max_{1\leq k\leq T}\sqrt{\frac{T}{2}}|D_T(k)| \Rightarrow \frac{\mathcal{M}_4^{1/2}}{\sqrt{2}\left(\sigma_\varepsilon^2 + p\lambda^2 + p\delta^2 + \sigma_\xi^2\right)}\sup_{\tau\in[0,1]}|W^*(\tau)|.$$

For the KL test, assume that $\widehat{\mathcal{M}}_{4,T}$ is a consistent estimator of $\mathcal{M}_4$. Standard HAC-type estimators render this property under the set of assumptions discussed. From Lemma 3.2 and the continuous mapping theorem it follows straightforwardly that,

$$\frac{\arg\max_{1\leq k\leq T}|G_T(k)|}{\sqrt{T}\widehat{\mathcal{M}}_{4,T}} = \frac{\arg\max_{1\leq k\leq T}\left|C_T(k) - \left(\frac{k}{T}\right)C_T(T)\right|}{\sqrt{T}\widehat{\mathcal{M}}_{4,T}} \Rightarrow \sup_{\tau\in[0,1]}|W^*(\tau)|.$$

This completes the proof.∎

Proof of Theorem 3.2.
Note that we can write $\widehat{k} = \arg\max_{1\leq k\leq T}\left|\left(\frac{k(T-k)}{T}\right)\frac{1}{k}\sum_{t=1}^{k}r_t^2 - \frac{1}{T-k}\sum_{t=k+1}^{T}r_t^2\right| \equiv \arg\max_{1\leq k\leq T}|R_{k,T}|$ where $R_{k,T}$ is defined implicitly. Next, note $E(R_{k,T}) = \Delta\kappa(1-\kappa^*)\mathbf{1}_{k\leq k^*} + \Delta\kappa^*(1-\kappa)\mathbf{1}_{k>k^*}$, and $E(R_{k^*,T}) = \Delta\kappa^*(1-\kappa^*)$, so $|E(R_{k^*,T})| - |E(R_k)| = |\Delta|(\kappa^* - \kappa)(1-\kappa)\mathbf{1}_{k\leq k^*} + |\Delta|\kappa^*(\kappa - \kappa^*)\mathbf{1}_{k>k^*}$ with $\mathbf{1}_{(\cdot)}$ being an indicator function. Hence $|E(R_{k^*,T})| - |E(R_k)| \geq |\Delta||\kappa^* - \kappa|\min\{\kappa^*, 1-\kappa^*\}$. Setting $\kappa = \widehat{\kappa}$, it can be shown that $|\Delta||\kappa^* - \widehat{\kappa}|\min\{\kappa^*, 1-\kappa^*\} \leq 2\max_{1\leq k\leq T}|R_{k,T} - E(R_{k,T})|$ or, equivalently,

$$|\widehat{\kappa} - \kappa^*| \leq \frac{1}{T}\arg\max_{1\leq k\leq T}\frac{4\sum_{t=1}^{k}|r_t^2 - E(r_t^2)|}{|\Delta|\min\{\kappa^*, 1-\kappa^*\}}.$$

Next, denote $\Psi_{k,T} = T^{-1}\arg\max_{1\leq k\leq T}\sum_{t=1}^{k}|r_t^2 - E(r_t^2)|$. For some $\epsilon > 0$, and Theorem 4.1 in Kokoszka and Leipus (2000), it follows that

$$\Pr(\Psi_{k,T} > \epsilon) \leq \frac{2}{\epsilon^2 T^2}\sum_{t=0}^{T-1}\sqrt{Var(r_{t+1}^2)\sum_{i,j=1}^{t}Cov(r_i^2, r_j^2)} + \frac{1}{\epsilon^2 T^2}\sum_{t=0}^{T-1}Var(r_{t+1}^2).$$

Under A2 and A4, $\text{Cov}(\varepsilon_i^2, \varepsilon_j^2) = O(\rho^{|t-j|})$, so $Cov(r_i^2, r_j^2)$ has finite upper bounds that decay exponentially. Also, $Cov(r_i^2, r_j^2) = Cov(\varepsilon_i^2, \varepsilon_j^2) + Var(Z_i)\mathbf{1}_{i=j} + Var(\xi_i)\mathbf{1}_{i=j}$. For $i = j$, $0 \leq Cov(r_i^2, r_j^2) \leq \pi$, with $\pi \equiv C_1^* + Var(Z_i) + Var(\xi_i) < \infty$, for some constant $0 < C_1^* \leq \sup_t E(\varepsilon_t^4)$, from Cauchy-Schwartz's inequality. Note for $i \neq j$, $Cov(r_i^2, r_j^2) = Cov(\varepsilon_i^2, \varepsilon_j^2)$ and, therefore, $0 \leq Cov(r_i^2, r_j^2) \leq C_1^*\rho^{|i-j|}$ with $0 \leq \rho < 1$ ruling the correlation pattern as a function of the specific model. Since $0 \leq Cov(r_i^2, r_j^2) \leq \pi\rho^{|i-j|}$ uniformly for $1 \leq i,j \leq T$, and denoting $\sigma_{\sup}^2 = \max_{1<t<T}Var(\varepsilon_t^2)$, we have $\Pr(\Psi_{k,T} > \epsilon) \leq \frac{K_1}{\epsilon^2\sqrt{T}} + \frac{\sigma_{\sup}^2}{\epsilon^2 T} \leq \frac{K_2}{\epsilon^2\sqrt{T}}$, where $K_1 = \frac{4}{3}\sqrt{\sigma_{\sup}^2\pi/(1-\rho)}$, and some constant

$K_2 > K_1 > 0$. Finally,

$$
\begin{aligned}
\Pr\left(|\widehat{\kappa} - \kappa^*| > \epsilon\right) &\leq \Pr\left(\Psi_{k,T} > \frac{\epsilon|\Delta|\min\{\kappa^*, 1-\kappa^*\}}{4}\right) \leq \frac{16K_2}{\epsilon^2\Delta^2\left(\min\{\kappa^*, 1-\kappa^*\}\right)^2\sqrt{T}} \\
&= \frac{C}{\epsilon^2\Delta^2\sqrt{T}}.
\end{aligned}
$$

This completes the proof.■

References

[1] Aggarwal R, Inclán C, Leal R. 1999. Volatility in emerging stock markets. *Journal of Financial and Quantitative Analysis* **34**: 33-55.

[2] Andreou E, Ghysels E. 2002. Detecting multiple breaks dynamics. *Journal of Applied Econometrics* **17**: 579-600.

[3] Andreou E, Ghysels E. 2004. The impact of sampling frequency and volatility estimators and change-point test. *Journal of Financial Econometrics* **2**: 290-318.

[4] Bai J. 1998. Estimation of multiple-regime regressions with least absolutes deviation. *Journal of Statistical Planning* **74**: 103-134.

[5] Chen G, Choi Y, Zhou Z. 2005. Nonparametric estimation of structural change points in volatility models for time series. *Journal of Econometrics* **16**: 79-114.

[6] Cuñado J, Gómez-Biscarri J, Pérez F. 2006. Changes in the dynamic behavior of emerging market volatility: Revisiting the effects of financial liberalization. *Emerging Markets Review* **7**: 261-278.

[7] Deng A, Perron P. 2008a. The limit distribution of the Cusum of squares test under general mixing conditions. *Econometric Theory* **24**: 809-822.

[8] Deng A, Perron P. 2008b. A non-local perspective on the power properties of the CUSUM and CUSUM of squares tests for structural change. *Journal of Econometrics* **142**: 212-240.

[9] Franses PH, Haldrup H. 1994. The effects of additive outliers on tests of unit Roots and cointegration. *Journal of Business and Economic Statistics* **12**: 471-478.

[10] Fiteni I. 2004. τ-estimators of regression models with structural change of unknown location. *Journal of Econometrics* **119**: 19-44.

[11] Inclán C, Tiao CJ. 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statiscal Association* **89**: 913-923.

[12] Kokoszka P, Leipus R. 2000. Change-point estimation in ARCH models. *Bernoulli* **6**: 513-539.

[13] McConnell M, Perez-Quirós G. 2000. Output fluctuations in the United States: what has changed since the early 1980s? *American Economic Review* **90**: 1464-76.

[14] Pagan AR, Schwert GW. 1990. Testing for covariance stationarity in stock market data. *Economics Letters* **33**, 165-170.

[15] Sensier M, and van Dijk D. 2004. Testing for volatility changes in US macroeconomic time series, *Review of Economics and Statistics* **86**: 833-839.

[16] van Dijk D, Franses PH, Lucas A. 1999. Testing for ARCH in the presence of additive outliers. *Journal of Applied Econometrics* **14**: 539-562.

[17] Vogelsang T. 1999. Two simple procedures for testing for a unit root when there are additive outliers. *Journal of Time Series Analysis* **20**: 237–52.

[18] White H. 2000. *Asymptotic theory for econometricians.* Academic Press.

A Figures

Figure 1. Empirical size of the Inclán-Tiao test (5% nominal size) with outlier-contaminated data.
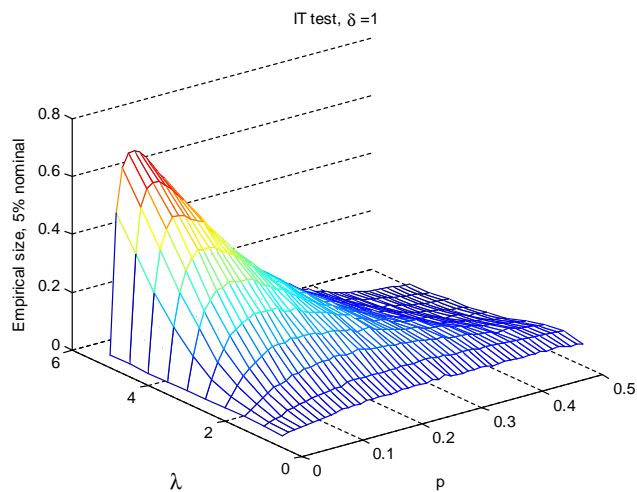
Note: The DGP is $r_t = \sigma \eta_t + D_t \left[\lambda + \delta v_t \right]$, $\eta_t, v_t \sim iid\mathcal{N}(0,1)$, $D_t = \{\pm 1, 0\}$ with probabilities $\{p/2, 1-p\}$. The results are based on 15,000 simulations for $T = 1000$ and $\delta = 1$. The test statistic is compared to the critical values from SSBB under the null of variance homogeneity. The experimental proportion of rejections are displayed on the vertical axis.

Figure 2. Empirical size of the Kokoszka-Leipus test (5% nominal size) with outlier-contaminated data.

KL test, δ =1
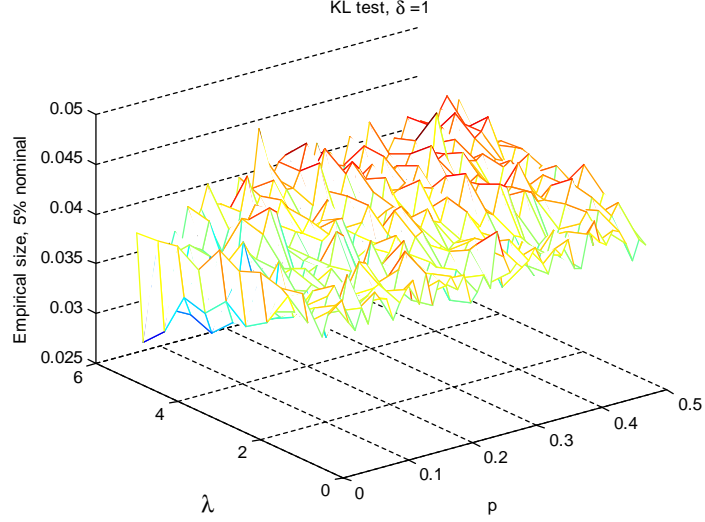
Note: The DGP is $r_t = \sigma\eta_t + D_t\left[\lambda + \delta v_t\right]$, $\eta_t, v_t \sim iid\mathcal{N}\left(0,1\right)$, $D_t = \{\pm1, 0\}$ with probabilities $\{p/2, 1-p\}$. The long-run variance of $r_t^2 - E\left(r_t^2\right)$ is computed using the Newey-West estimator with Bartlett kernel and bandwidth $h = \left[4\left(100/T\right)^{2/9}\right]$. The results are based on 15,000 simulations for $T = 1000$ and $\delta = 1$. The test statistic is compared to the critical values from SSBB under the null of variance homogeneity. The experimental proportion of rejections are displayed on the vertical axis.

Figure 3. Empirical size of the Inclán-Tiao test (5% nominal size) with GARCH [GARCH2] errors and outliers.

IT test, δ =1
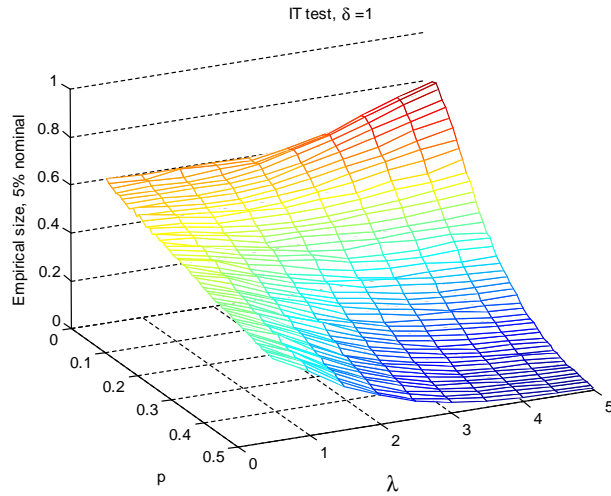
Note: See caption under Figure 1, but considering in this case that the conditional volatility σ_t^2 follows a GARCH(1,1) process with parameters $(\alpha,\beta) = (0.1, 0.8)$.

Figure 4. Empirical size of the Kokoszka-Leipus test (5% nominal size) with GARCH errors [GARCH1] and outliers.
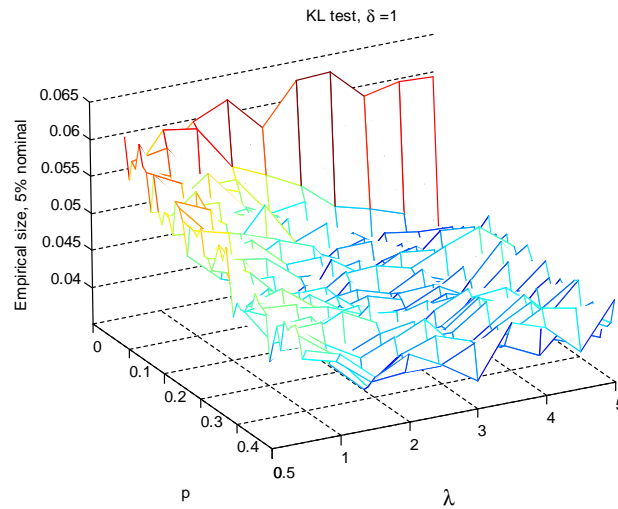
KL test, δ =1

Note: See caption under Figure 2, but considering in this case that the conditional volatility σ_t^2 follows a GARCH(1,1) process with parameters $(\alpha, \beta) = (0.1, 0.5)$.

Figure 5. Empirical size of the Kokoszka-Leipus test (5% nominal size) with GARCH errors [GARCH2] and outliers.
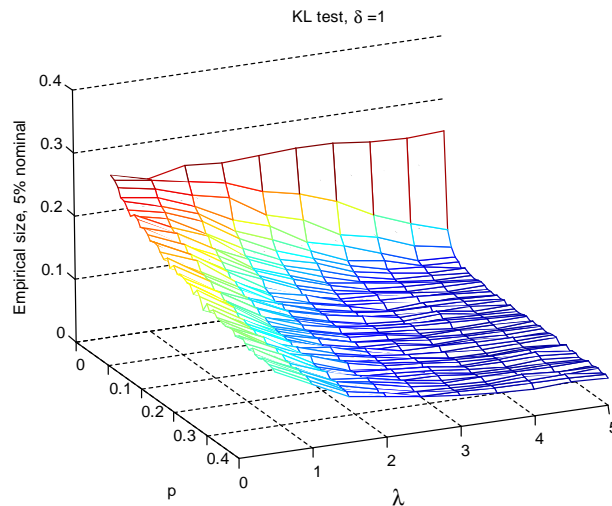
KL test, δ =1

Note: See caption under Figure 2, but considering in this case that the conditional volatility σ_t^2 follows a GARCH(1,1) process with parameters $(\alpha, \beta) = (0.1, 0.8)$.

A Tables

TABLE 1. Average estimation ($E(\widehat{\kappa})$) and standard error (s.e. x 100) of $\widehat{\kappa}$ based on r_t^2.

| | iid κ^* | | | | | | GARCH1 κ^* | | | | | | GARCH2 κ^* | | | | | |
| | 0.25 | | 0.50 | | 0.75 | | 0.25 | | 0.50 | | 0.75 | | 0.25 | | 0.50 | | 0.75 | |
(λ,p)	$E(\widehat{\kappa})$	s.e.	$E(\widehat{\kappa})$	s.e.	$E(\widehat{\kappa})$	s.e.	$E(\widehat{\kappa})$	s.e.	$E(\widehat{\kappa})$	s.e.	$E(\widehat{\kappa})$	s.e.	$E(\widehat{\kappa})$	s.e.	$E(\widehat{\kappa})$	s.e.	$E(\widehat{\kappa})$	s.e.
(0,0)	0.289	0.05	0.509	0.02	0.745	0.02	0.306	0.07	0.514	0.03	0.742	0.03	0.358	0.12	0.530	0.06	0.728	0.08
(0,0,10)	0.293	0.06	0.509	0.02	0.743	0.03	0.309	0.08	0.515	0.03	0.740	0.04	0.361	0.12	0.530	0.06	0.728	0.08
(0,0,25)	0.297	0.07	0.510	0.02	0.739	0.04	0.312	0.08	0.515	0.03	0.736	0.04	0.362	0.13	0.530	0.06	0.724	0.09
(0,0,50)	0.304	0.08	0.512	0.03	0.734	0.04	0.319	0.09	0.519	0.04	0.733	0.05	0.367	0.13	0.530	0.06	0.720	0.09
(1,0)	0.289	0.05	0.509	0.02	0.745	0.02	0.307	0.07	0.514	0.03	0.742	0.03	0.358	0.12	0.530	0.05	0.730	0.08
(1,0,10)	0.297	0.06	0.501	0.02	0.739	0.03	0.314	0.08	0.515	0.03	0.736	0.04	0.361	0.12	0.530	0.06	0.723	0.09
(1,0,25)	0.306	0.08	0.510	0.03	0.734	0.05	0.322	0.09	0.516	0.04	0.729	0.06	0.367	0.13	0.530	0.06	0.717	0.10
(1,0,50)	0.320	0.10	0.513	0.04	0.721	0.07	0.333	0.11	0.518	0.05	0.719	0.08	0.375	0.14	0.532	0.07	0.707	0.11
(2.5,0)	0.289	0.05	0.509	0.02	0.746	0.02	0.307	0.08	0.514	0.03	0.742	0.03	0.357	0.12	0.528	0.06	0.729	0.08
(2.5,0,10)	0.320	0.10	0.510	0.05	0.714	0.08	0.333	0.11	0.515	0.05	0.710	0.09	0.378	0.14	0.531	0.08	0.701	0.12
(2.5,0,25)	0.351	0.14	0.514	0.08	0.683	0.12	0.367	0.14	0.517	0.08	0.682	0.13	0.394	0.16	0.529	0.10	0.670	0.15
(2.5,0,50)	0.391	0.17	0.518	0.11	0.655	0.16	0.396	0.17	0.524	0.11	0.656	0.16	0.420	0.17	0.536	0.12	0.647	0.18
(5,0)	0.289	0.06	0.509	0.02	0.746	0.02	0.307	0.08	0.514	0.03	0.742	0.03	0.358	0.12	0.530	0.06	0.729	0.08
(5,0,10)	0.412	0.19	0.506	0.14	0.603	0.19	0.415	0.19	0.509	0.14	0.604	0.19	0.431	0.19	0.519	0.15	0.601	0.20
(5,0,25)	0.455	0.20	0.514	0.17	0.566	0.21	0.459	0.20	0.513	0.17	0.569	0.21	0.465	0.20	0.522	0.18	0.567	0.21
(5,0,50)	0.477	0.21	0.523	0.18	0.556	0.22	0.485	0.21	0.523	0.19	0.561	0.22	0.488	0.21	0.526	0.19	0.556	0.22

Note: Average value of the break-fraction estimator $\widehat{\kappa}$ and standard error (x100), with $\widehat{\kappa} = k/T$, and $\widehat{k} : \max_{1\leq k\leq T} \left| \sum_{t=1}^{k} r_t^2 - k/T \sum_{t=1}^{T} r_t^2 \right|$. The DGP is $r_t = \sigma_t \eta_t + \mathcal{B}_t [\lambda + v_t]$, with $\eta_t, v_t \sim iidN(0,1)$, $\mathcal{B}_t = \{-1,1,0\}$ with probabilities $\{p/2, p/2, 1-p\}$. For $t < [\kappa^*T]$, $E(\sigma_t \eta_t) = 1$ and $E(\sigma_t \eta_t) = 1.5$ otherwise. The volatility process σ_t follows an iid sequence, and a GARCH(1,1) model with parameters $(\alpha, \beta) = (0.1, 0.5)$ [GARCH1], and $(\alpha, \beta) = (0.1, 0.8)$ [GARCH2].

TABLE 2. Testing for multiple change-points in the volatility of monthly US market index returns.

Panel A: Absolute values, $|r_t|$									
Inclan-Tiao					*Kokoszca-Leipus*				
0%	0.5%	1.0%	2.0%	2.5%	0%	0.5%	1.0%	2.0%	2.5%
1893:04	1893:04	1928:10	1901:07	1893:04		1928:10	1928:10		1927:09
1893:12	1893:12	1939:09	1927:06	1904:04		1939:09	1939:09		1939:08
1929:05	1928:10	1973:09	1939:08	1927:09		1973:09	1973:09		
1931:08	1933:11	1975:01	1973:09	1939:08					
1933:06	1937:08		1987:01						
1938:02	1939:09								
1938:06	1973:09								
1940:04	1975:01								
1940:06	1987:09								
1973:09									
1975:01									
1987:09									

Panel B: Squared Values, r_t^2									
Inclan-Tiao					*Kokoszca-Leipus*				
0%	0.5%	1.0%	2.0%	2.5%	0%	0.5%	1.0%	2.0%	2.5%
1893:06	1885:09	1896:06	1885:09	1885:09					1928:02
1893:08	1885:10	1901:04	1885:10	1885:10					1934:07
1907:02	1891:07	1907:09	1891:07	1893:04					
1907:09	1893:06								
1928:10									
1940:05									
1973:09									
1973:11									
1974:08									

Note: The date of the break is estimated under the IT and KL test procedures at the 5% nominal significance level and given the trimming fractions presented in the columns (0%, 0.5%, 1.0%, 2.0%, 2.5%). For instance, 0.5% indicates that the tests are applied after removal of the 0.5% most extreme observations in the sample.

WORKING PAPERS

2008