A comparative Study of Volatility Breaks

Claudia Grote*and Philip Bertram

2013

Institute of Statistics,
Faculty of Economics and Management,
Leibniz University Hannover,
D-30167 Hannover, Germany.

Abstract

In this paper we evaluate the performance of several structural break tests under various DGPs. Concretely we look at size and power properties of CUSUM based, LM and Wald volatility break tests. In a simulation study we derive the properties of the tests under shifts in the unconditional and conditional variance as well as for smooth shifts in the volatility process. Our results indicate that Wald tests have more power of detecting a change in the volatility than CUSUM and LM tests. This, however, goes along with the disadvantage of being slightly oversized. We further show that with huge outliers in the data the tests may exhibit non-monotonic power functions as the long-run variance of the squared return process is no longer finite. In an empirical example we determine the number and time of volatility breaks considering four equity and three exchange rate series. We find that in some situations the outcomes of the tests may vary substantially. Further we find fewer volatility breaks in the currency series than in the equity series.

JEL codes: C22, C52, C53

Keywords: structural breaks, variance shifts, non-monotonic power

*Corresponding author: phone: +49 511 762-3082, fax: +49- 511- 762-3923,
e-mail: grote@statistik.uni-hannover.de.

1 Introduction

During the last couple of decades, a great deal of attention has been drawn to volatility shifts and their impact on time series in the context of financial markets. Ever since the seminal articles of Diebold [1986] and Lamoureux and Lastrapes [1990] stylized facts of volatility such as long-range dependence or IGARCH effects are regarded as being caused by structural changes in volatility. There exists evidence that many time series suffer from occasional structural breaks in the conditional as well as the unconditional volatility, compare e.g. Andreou and Ghysels [2002], Sensier and van Dijk [2004]. Consequently there are several proposals in the literature for incorporating structural changes in volatility into GARCH models, cf. Engle and Rangel [2008], Engle et al. [2013] or Amado and Teräsvirta [2013] among others. Hence, testing for volatility constancy marks an important task in terms of model selection and forecasting purposes. Break detection plays also an essential role for e.g. financial decision-making like the pricing of derivatives and portfolio risk management. Since the implicit assumption of a stable underlying GARCH process is often confuted by sudden large shocks the unconditional volatility of exchange rate returns can in turn be effected, c.f. Rapach and Strauss [2008]. Another strand of literature where break detection is of great importance is the influence of volatility on causality, c.f. D.O. van Dijk and Sensier [2005]. The most widely employed testing procedures for treating the aforementioned issues in the field of volatility breaks are commonly CUSUM, LM and Wald tests.

Empirical applications considering structural breaks in GARCH processes applied to the latter tests in particular have i.a. been provided by Andreou and Ghysels [2002], Sensier and van Dijk [2004] and Xu [2013b]. The former authors have compared CUSUM and least-squares volatility break tests concerning their size and power performance applied to GARCH processes. By additionally considering shifts in the unconditional variance process Xu [2013b] looked at a broader range of data generating processes (DGPs).

Our paper deals with the most frequently used volatility break tests and compares them over a broad range of different DGPs. Thereby, we look at switches in the unconditional as well as the conditional volatility, whereat the underlying DGP is either exposed to single or double shifting or can alternatively exhibit a smooth and steady increase in the magnitude of the volatility break. The comparison is done via an extensive simulation study at which we apply seven different tests. Further, we elaborate that for certain parameter constellations the tests may suffer from non-monotonic power functions, provided that the data contains large outliers. This results from the fact that the long-run variance of the squared process may no longer be finite causing the power to drop once the finite kurtosis condition is no longer fulfilled. By estimating the density of the break point estimators we assess the correctness of the estimation and can further confirm the findings of the simulation study.

In the empirical application we analyze four equity series and three exchange rate series. By estimating the number and timing of the breakpoints we see that the outcome of the tests can indeed differ substantially for different parameter constellations. Following we carry out the tests for each series in a rolling window allowing us to derive and compare the distribution of the p-values of the tests. Here we find that for currencies fewer breaks in volatility are found compared with the equity series.

The paper is organized as follows. Section 2 provides a short introduction of the insinuated testing procedures. In section 3 the simulation study is presented and the results are discussed. Section 4 then contains the empirical example while section 5 concludes.

2 Volatility Break Tests

Let $\{\eta_t\}_{t=1}^{T}$ denote a mean-zero stochastic process with index t and T the time horizon. To test for a possible break in the volatility of the process CUSUM, Lagrange multiplier (LM) and Wald tests have been applied in the literature.

The CUSUM-of-squares test originally introduced by Brown et al. [1975] is given by

$$CUSQ = \max_{0\leq\tau\leq 1} \left| \sum_{t=1}^{[\tau T]} \eta_t^2 - \sum_{t=1}^{T} \eta_t^2 \right| / \sqrt{T\Theta},$$

where $[\tau T]$ describes the break point that occurs at the time, with τ denoting the percentaged breakpoint $\tau \in [0,1]$, and Θ being the long-run variance (LRV) of the squared process. Under the Null of no volatility breaks $CUSQ \to \sup_{\tau\in[0,1]} |BB(\tau)|$ where $BB(\tau) = W(\tau) - \tau W(1)$ with $W(\tau)$ and $BB(\tau)$ being defined as a unit Wiener process on [0,1] and a Brownian Bridge in dependence of τ, respectively. Simulating $BB(\tau)$ results in critical values of 1.224(10%), 1.358(5%) and 1.628(1%) and can be found in. Deng and Perron [2008] specify $\{\eta_t\}_{t=1}^{T}$ to be α-mixing and formulate an estimator for the LRV under the Null. Earlier Inclan and Tiao [1994] looked at variance breaks for normal *iid* data. Thereby the LRV of the CUSQ simplifies to $\Theta = 2$. As the version of Inclan and Tiao [1994] is heavily oversized for dependent data (cf. Andreou and Ghysels [2002]) we merely look at the Deng and Perron [2008] version (henceforth DP) of the test in the upcoming analysis.

Alternatively the LM test can be utilized. The test statistic is given by

$$LM(\tau) = (SSR_0 - SSR(\tau)) / \sqrt{\Theta} \tag{1}$$

where SSR_0 denotes the sum of squared residuals of the simple mean shift model

$$\eta_t^2 - \sum_{t=1}^{T} \eta_t^2 / T = \varrho \mathbb{1}_{(t\geq[\tau T])} + \upsilon_t \;\text{ with }\; \upsilon_t \overset{iid}{\sim} \mathcal{N}(0,1).$$

Thereby ϱ depicts the mean shift parameter giving the break size and $\mathbb{1}$ depicts the indicator function leading to the Null of $H_0 : \varrho = 0$ while $SSR(\tau)$ is defined as the sum of squared residuals under the alternative of a break in the process at time $[\tau T]$.

The Wald test ist then specified as

$$Wald(\tau) = (SSR_0 - SSR(\tau)) / \sqrt{\Theta(\tau)}, \tag{2}$$

in which Θ is determined under the alternative, i.e. $\Theta(\tau)$.

In our analysis we consider supremum, mean and exponential versions of (1) and (2). The functionals of the tests $\mathcal{J} = \{LM, Wald\}$ are then defined as:

i) sup -$\mathcal{J} = \sup \mathcal{J}(\tau)$

ii) mean -$\mathcal{J} = \int \mathcal{J}(\tau)$

iii) exp -$\mathcal{J} = \log\left(\int \exp\left[\frac{1}{2}\mathcal{J}(\tau)\right]\right)$,

where the integral is defined over τ with $\tau \in \tau_\varepsilon$ where $\tau_\varepsilon = \{\tau : \tau \geq \varepsilon,\ \tau \leq 1-\varepsilon\}$. ε describes the user-chosen truncation parameter concerning the interval in which a break point is tested for. Throughout the analysis we set $\varepsilon = 0.15$, following Andrews [1991] here. The critical values for i) are taken from Andrews [1993] and those for ii) - iii) can be found in Andrews and Ploberger [1994].

In all three tests $\Theta = \gamma_0 + 2\sum_{r=1}^{\infty}\gamma_r$ marks the long-run variance (LRV) of η_t with $\gamma_r = E(\eta_t^2 - \sigma^2)(\eta_{t-r}^2 - \sigma^2)$ and $\sigma^2 = E(\eta_t^2)$ for $r = 0, 1, \ldots, T-1$. An estimator of Θ is given by $\hat{\Theta} = \hat{\gamma}_0 + 2\sum_{r=1}^{T-1} k(r/m)\hat{\gamma}_r$ with $\hat{\gamma}_r = T^{-1}\sum_{t=r+1}^{T}(\eta_t^2 - \hat{\sigma}^2)(\eta_{t-r}^2 - \hat{\sigma}^2)$ and $\hat{\sigma}^2 = \sum_{t=1}^{T}\eta_t^2/T$. Like many others we specify $k(\cdot)$ as the Bartlett kernel while the bandwidth m is determined conducting the data-dependent method with an AR(1) approximation proposed by Andrews [1991]. Hence, the estimated bandwidth $\hat{m}$ equals

$$\hat{m} = 4\hat{\rho}^2/(1-\hat{\rho}^2)^2 \tag{3}$$

where $\hat{\rho}$ denotes the OLS estimate from a regression of $\hat{\eta}_t^2$ on $\hat{\eta}_{t-1}^2$.

3 Monte Carlo Study

3.1 Simulation setup

By applying the presented tests to DGPs that underly a break in either the conditional or the unconditional volatility we want to issue potential pitfalls by means of size and power properties. Here we distinguish between three different process types on behalf of the shift type: single shifting (I), double shifting (II) and smooth transition (III). Following Xu [2013b] the DGPs have the form $\eta_t = \sigma_t \epsilon_t$ and are composed of a conditional variance term ϵ_t and an unconditional variance term $\sigma_t^2 = \sigma^2(t/T)$ with $\sigma^2(s)$ being defined on $s \in (0,1]$, i.e.

$$\epsilon_t = h_t\xi_t,\ h_t^2 = \mu + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}^2,\ \xi_t \overset{iid}{\sim} \mathcal{N}(0,1)$$
$$\sigma^2(s) = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\cdot\{\text{I, II, III}\} \ \text{ with } \ \delta \equiv \frac{\sigma_1}{\sigma_0}.$$

By construction the DGPs depend on the parameters τ, δ and T denoting breakpoint, break size and sample size, respectively. Throughout the analysis σ_0^2 is set to equal one and h_t forms a simple $GARCH(1,1)$ process with ξ_t being iid normal.

As soon as the process breaks, the unconditional variance switches to one of the three process types {I,II,III}. The switches are incorporated into the processes in dependence of the design of the variance shift as follows:

$$\text{single shifts (I): } \mathbb{1}_{\{s\geq\tau\}},\ \tau\in\{0.3,0.5,0.7\}$$
$$\text{double shifts (II): } \mathbb{1}_{\{\tau\leq s\leq\tau+0.2\}},\ \tau\in\{0.3,0.5,0.7\}$$
$$\text{smooth transition (III): } s.$$

The DGPs of the type (I) or (II) are hence exposed to single or double shifting, incorporated via the indicator function $\mathbb{1}$. In contrast to type (I), processes of type (II) switch from σ_0^2 to σ_1^2 at $\lfloor\tau T\rfloor$, stay on for $\lfloor 0.2T\rfloor$, and switch back to the initial variance process, cf. Xu [2013b]. For a better overview a summary of the different DGPs and their properties is displayed in Figure 1.

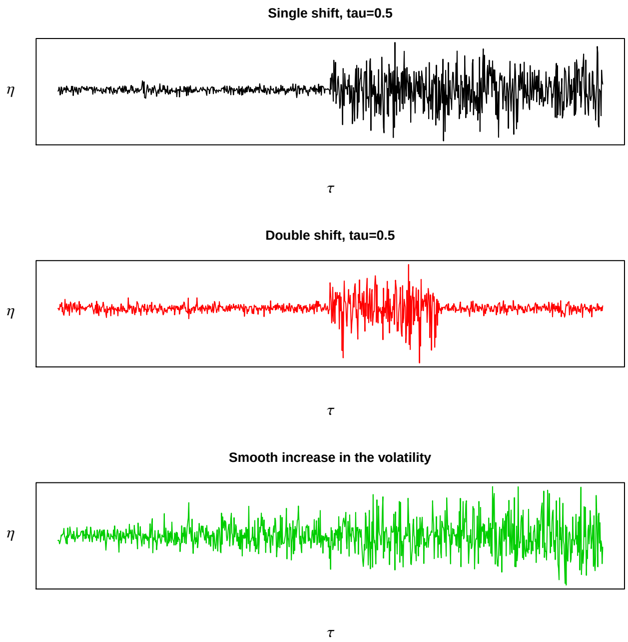

Figure 1: displays exemplary the inflation of the volatility over time for the different process types for $T = 1000$ with $\tau = 0.5$.

The break itself is then incorporated via the indicator function $\mathbb{1}$ as soon as t exceeds the predetermined τ. Whenever the break point $\lfloor\tau T\rfloor$ is reached, the magnitude of the break, δ, switches from $\delta = 1$ under the Null to $\delta \in \{1.1, 1.2, 1.3, 1.5, 2\}$ under the alternative and, thus, causes the volatility shift. The considered samples sizes are $T = 200, 500, 1000$. All of these DGPs are subject to a simple shift at certain breakpoints occurring at the smallest integer of $\lfloor\tau T\rfloor$. Thus, the break is defined to happen either after 30%, 50% or 70% of the time, meaning $\tau \in \{0.3, 0.5, 0.7\}$.

Within the class of single shift processes we use four different DGPs. DGP1 undergoes a switch in the unconditional volatility, while DGPs 2-4 are exposed to a break in the conditional volatility. The DGPs 2-4 are based on different specifications of the GARCH parameters μ, α and β. Following Hillebrand [2005] the values for $(\mu, \alpha, \beta) = (2\cdot 10^{-5}, 0.1, 0.4)$ are specified to equal an annualized volatility of $\sigma = \sqrt{250\mu/(1-\alpha-\beta)} = 0.1$. Hence, the DGPs 2-4 jump from the bold type values in (4) under the null hypothesis to the

	Break points	Single shift	Two Shifts	Smooth transition
DGP 1	(I) single shift	Simple break at τT		
DGP 2		$\boldsymbol{\mu}$ =(**2**, 2.42, 2.88, 3.38, 4.5, 8)$\cdot 10^{-05}$		
DGP 3		$\boldsymbol{\alpha}$ =**0.1**,.19, .25, .30, .378, .48		
DGP 4		$\boldsymbol{\beta}$ =**0.4**,.487, .553, .604, .678, .78		
DGP 5	(II) double shift		breaks for $0.2\tau T$	
DGP 6			$\xi_t \overset{iid}{\sim} \sqrt{0.6}t(5)$	
DGP 7	(III) smooth transition			$s = \tau \in [0,1]$
DGP 8				$s = \sqrt{\tau} \in [0,1]$
DGP 9				$s = \tau^2 \in [0,1]$

Table 1: graphs an overview of the DPGs in the simulation study.

following values in accordance with the magnitude of $\delta = \mathbf{1}, 1.1, 1.2, 1.3, 1.5, 2$:

$$\begin{aligned} &\textit{in } \text{DGP2 } \mu \text{ varys:} && \boldsymbol{\mu} = \{(\mathbf{2}, 2.42, 2.88, 3.38, 4.5, 8) \cdot 10^{-05}\} \\ &\textit{in } \text{DGP3 } \alpha \text{ varys:} && \boldsymbol{\alpha} = \{\mathbf{0.1}, .19, .25, .30, .378, .48\} \\ &\textit{in } \text{DGP4 } \beta \text{ varys:} && \boldsymbol{\beta} = \{\mathbf{0.4}, .487, .553, .604, .678, .78\}. \end{aligned} \tag{4}$$

Additionally, we consider DGPs that are exposed to double shifting (DGPs 5-6) and processes that undergo a smoothly increasing expansion in the unconditional volatility over time. While DGP 5 is exposed to simple double shifting, DGP 6 has additionally heavy tails, where the error term $\xi_t \overset{iid}{\sim} \sqrt{0.6}t(5)$, depicting a t-distribution normalized to mean zero and standard deviation one. DGPs 7-9 exhibit a smooth transition in the volatility, depending on the transition parameter s. Proportional to the time period the DGPs 7-9 evolve with $s = \tilde{s}, \sqrt{\tilde{s}}, \tilde{s}^2 \in [0,1]$. For a better visualization Figure 1 displays the inflation of the volatility over time for the different process types, respectively.

Apart from DGP 7 and DGP 8, which are taken from Cavaliere and Taylor [2007] and Xu [2008] the remaining DGPs are adopted from Xu [2013b]. Finally, to review the different DGPs Table 1 sums up their properties.

3.2 Simulation results

As the results of our study contain 5 dimensions (sample sizes, breakpoints, break sizes, DGPs and tests) we focus on the most striking patterns in the discussion of our results. That is, we only present the results for breaks in the middle of the sample, i.e. $\tau = 0.5$ and for small ($T = 200$) and large ($T = 1{,}000$) samples.[1]

To get a first idea of the behavior of the tests we look at the power of the tests concerning the different DGPs. As all tests behave qualitatively in the same way regarding the DGPs we focus on the power results for the DP test for illustrative purposes. Figure 2 plots the power in small and large samples for $\tau = 0.5$ for all DGPs.

As one can clearly see, the power for the single shift DGPs in small samples is naturally higher than for two shift DGPs. Furthermore it is striking that for the smoothly increasing volatility processes the power is substantially lower than for (single) discrete

[1] Further results are available upon request. The results for $\tau = 0.3$ and $\tau = 0.7$ are qualitatively not different from $\tau = 0.5$.

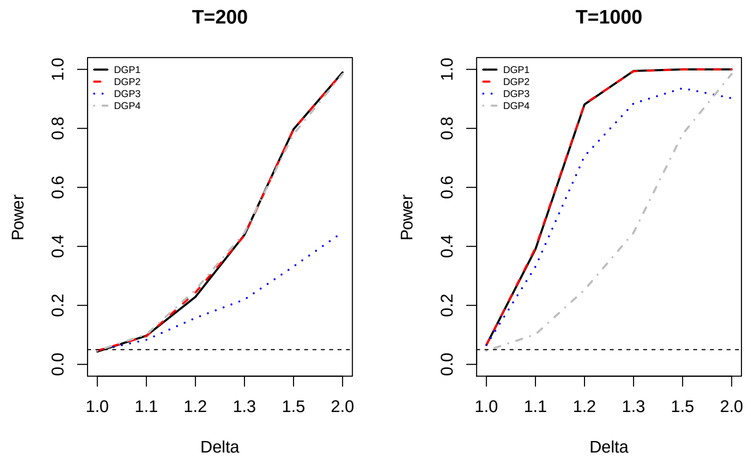

Figure 2: displays the power of the DP test for $T = 200$ (left) and $T = 1000$ (right) with $\tau = 0.5$ over all DGPs.

shifts in both, small and large, samples (compare the upper and the lower first graphs in Figure 2). In large samples the power converges to 1 for nearly all DGPs but DGP 3. It is quite interesting that the DGPs 1, 2 and 4 behave very alike and attain good power results especially when the magnitude of the break is very distinct. Hence, a first observable phenomenon is, that there is a negligible difference between the unconditional and conditional break in the variance, if we look at the DGP 1 compared to DGP 2 or 4. Secondly we can state that the process with the break in the ARCH parameter, DGP 3, even suffers from a power drop in large samples as soon as the magnitude of the break exaggerates $\delta > 1.5$. This non-monotonic power in DGP 3 can also be observed for all types of the least-squares tests not pictured here. Tables 2 and 3 then return the results for all tests and DGPs for $T = 200$ and $T = 1,000$.[2]

Turning to the small sample results (cf. Table 2) first we identify the expWald tests to exhibit the highest power of all tests for nearly all DGPs. However, we also observe the expWald and the other two Wald tests to be slightly oversized - a fact that has already been pointed out regarding mean shift tests by other authors such as e.g. Kejriwal [2009]. Consequently the Wald versions of the tests attain always higher power than their respective LM counterparts. Comparing the three tests implying a breakpoint estimator, i.e. the DP, the supLM and the supWald test, one can say that in terms of power supWald ≻ DP ≻ supLM in small samples. In accordance with Figure 2 we state very low power for all tests concerning the double shift DGPs 5 and 6. Also the power for the smoothly inflating variance-processes, DGPs 7-9, is rather low.

Rather surprising is the fact that a switch in the ARCH parameter (DGP3) is clearly less often detected than a break in the GARCH parameter. This may be caused by outliers resulting from a large ARCH parameter. In such cases with large outliers in the data the volatility break tests may no longer be robust to large volatility switches as the long-run variance might no longer be finite. These considerations are further dealt

[2]Results for $T = 500$ can be found in Table 11 in the appendix.

with in section 3.3. Apart from this the difference between switches in the unconditional (DGP1) and conditional (DGPs 2-4) variance process does not play a role for the power of the tests.

Altogether the tests perform very akin in small samples which is due to the fact, that the LM and Wald test both are based on least squares and resemble each other in their procedures as well as the CUSQ.

	DGP 1							DGP 2							DGP 3						
		LM			Wald				LM			Wald				LM			Wald		
Σ	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp
1	4.34	4.84	5.74	5.90	7.10	6.80	7.70	4.64	5.46	6.04	6.56	6.30	5.60	7.10	4.64	5.46	6.04	6.56	6.30	5.60	7.10
1.1	9.72	8.54	12.82	12.02	11.70	14.20	14.30	9.56	8.88	12.34	11.86	13.10	13.60	14.50	8.34	8.94	11.24	11.10	13.10	13.30	14.30
1.2	22.88	19.34	28.08	26.14	25.50	29.90	31.50	24.36	20.90	29.10	27.58	28.10	31.80	32.70	15.74	15.58	20.50	20.06	21.90	23.00	25.60
1.3	43.92	37.40	48.90	47.28	46.60	52.90	53.90	43.70	36.60	49.02	47.08	48.90	55.70	55.80	21.98	20.74	27.96	27.26	32.00	33.40	36.10
1.5	79.70	71.30	83.18	82.12	79.40	83.50	85.30	79.62	72.22	82.58	81.62	82.40	87.20	87.20	33.18	31.04	41.14	39.92	42.60	47.10	48.60
2	98.98	97.72	99.42	99.46	99.70	99.90	100.0	98.90	98.12	99.30	99.34	99.60	99.80	99.80	44.80	42.02	52.86	52.16	55.90	60.80	62.40
	DGP 4							DGP 5							DGP 6						
		LM			Wald				LM			Wald				LM			Wald		
Σ	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp
1	4.64	5.46	6.04	6.56	6.30	5.60	7.10	4.34	4.84	5.74	5.90	7.10	6.80	7.70	3.46	4.56	5.42	5.68	7.50	7.90	8.00
1.1	10.16	9.48	12.48	12.46	13.70	14.20	15.40	4.46	4.10	5.48	5.46	6.10	4.40	6.00	3.08	4.38	5.52	5.26	5.60	5.70	6.50
1.2	25.26	22.50	29.60	28.48	30.00	32.90	34.00	5.52	3.86	6.00	6.00	6.70	6.60	7.60	2.66	2.84	3.78	3.86	5.90	5.50	5.90
1.3	44.64	38.00	48.64	47.56	50.40	55.90	56.30	7.86	4.56	7.14	7.90	9.30	9.90	11.80	3.68	3.06	4.54	4.32	5.70	5.60	6.80
1.5	78.04	71.72	80.46	79.64	81.50	85.10	85.60	11.98	6.94	10.10	11.20	11.40	12.30	14.50	5.60	2.96	5.40	5.64	5.10	6.30	7.50
2	98.44	97.18	98.30	98.64	99.90	99.40	99.40	26.82	14.22	26.88	24.68	24.70	33.80	34.00	12.32	5.74	10.88	10.72	12.40	15.10	17.50
	DGP 7							DGP 8							DGP 9						
		LM			Wald				LM			Wald				LM			Wald		
Σ	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp
1	4.58	5.26	6.04	6.36	7.70	7.70	9.00	4.58	5.26	6.04	6.36	7.70	7.70	9.00	4.58	5.26	6.04	6.36	7.70	7.70	9.00
1.1	6.42	7.04	8.50	8.64	10.80	10.30	11.60	5.64	5.88	7.64	7.52	9.00	9.90	9.70	6.72	7.66	8.88	9.28	11.50	10.90	12.20
1.2	11.06	10.92	15.40	14.28	16.80	18.30	19.10	8.02	8.48	11.22	10.92	13.10	13.60	14.50	12.02	13.00	16.34	15.76	19.40	19.50	20.10
1.3	17.92	17.66	25.06	23.42	25.40	28.80	29.20	11.68	11.34	16.12	15.04	17.10	18.60	20.20	20.60	21.84	28.22	27.10	29.80	31.40	33.50
1.5	35.86	34.96	47.96	45.04	44.70	51.50	51.10	21.26	20.14	30.54	27.74	28.50	35.90	34.40	42.78	44.18	54.78	52.58	53.60	58.70	58.20
2	75.28	72.54	87.68	84.00	82.20	90.40	89.20	45.30	42.50	61.06	56.12	55.40	66.80	64.60	86.04	85.76	93.56	91.84	92.50	95.60	95.30

Table 2: reports power results for all DGPs according to the DP, LM- and Wald-type tests for $\tau = 0.5$, $\epsilon = 0.15$ and $T = 200$.

	DGP 1							DGP 2							DGP 3						
		LM			Wald				LM			Wald				LM			Wald		
Σ	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp
1	6.60	6.66	6.78	7.02	8.00	8.20	8.00	6.56	6.54	6.48	7.22	7.50	5.80	7.30	6.56	6.54	6.48	7.22	7.50	5.80	7.30
1.1	39.04	34.44	39.46	38.66	36.80	40.00	40.30	39.44	34.22	39.82	39.18	37.60	41.60	41.80	33.12	30.08	34.16	34.00	33.50	36.50	37.60
1.2	88.12	83.96	87.42	87.14	83.90	87.30	87.20	88.28	83.70	87.20	87.24	85.00	87.30	87.80	70.60	64.96	70.92	70.64	67.20	71.60	71.70
1.3	99.46	99.08	99.28	99.40	99.30	99.20	99.50	99.36	98.98	99.30	99.38	99.50	99.80	99.90	88.44	84.94	88.82	88.76	86.50	89.10	89.70
1.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	93.62	92.00	94.62	94.82	93.60	96.20	95.70
2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	90.22	90.10	92.60	92.98	91.70	92.90	93.20

	DGP 4							DGP 5							DGP 6						
		LM			Wald				LM			Wald				LM			Wald		
Σ	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp
1	6.56	6.54	6.48	7.22	7.50	5.80	7.30	6.60	6.66	6.78	7.02	8.00	8.20	8.00	4.84	6.20	5.96	6.64	6.60	6.00	6.30
1.1	39.60	34.70	39.98	39.36	38.00	41.90	42.10	11.64	9.10	9.62	10.64	9.70	10.10	11.70	5.08	4.54	5.46	5.62	4.50	5.30	5.40
1.2	87.58	83.14	86.28	86.56	84.10	86.60	87.20	26.10	20.94	19.00	23.36	23.10	20.20	25.20	9.36	6.58	8.04	9.04	6.50	9.00	9.30
1.3	99.22	98.72	99.08	99.18	99.40	99.60	99.90	50.68	42.88	42.56	47.54	46.50	42.50	49.50	18.74	13.16	15.48	17.46	13.40	14.90	16.30
1.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	92.20	88.04	91.44	92.30	90.60	92.80	93.80	44.68	34.76	41.00	42.92	34.20	38.70	41.90
2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.94	99.88	99.98	99.94	99.90	100.0	100.0	89.08	83.92	90.06	89.66	85.70	89.30	90.10

	DGP 7							DGP 8							DGP 9						
		LM			Wald				LM			Wald				LM			Wald		
Σ	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp
1	6.38	7.26	6.60	7.46	6.70	6.20	6.20	6.38	7.26	6.66	7.46	6.70	6.20	6.20	6.38	7.26	6.66	7.46	6.70	6.20	6.20
1.1	18.54	17.28	21.10	20.22	19.10	20.80	20.90	13.50	13.20	15.38	15.26	15.00	15.60	15.50	19.52	19.46	21.80	21.84	21.10	21.80	22.30
1.2	46.52	44.60	51.98	49.74	49.20	55.10	53.40	30.50	29.64	35.06	34.16	32.90	37.50	37.10	50.46	49.64	55.38	54.54	53.80	59.30	58.60
1.3	75.86	74.32	82.30	80.50	75.40	82.60	81.50	52.40	51.12	59.68	57.36	51.80	59.50	58.00	81.54	80.56	85.74	84.60	81.70	85.70	84.90
1.5	97.98	97.62	98.98	98.70	98.10	99.40	99.00	85.18	84.76	90.22	89.38	87.70	91.50	91.20	99.12	99.14	99.46	99.46	99.50	99.80	99.80
2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.62	99.68	99.92	99.92	99.80	99.90	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 3: reports power results for all DGPs according to the DP, LM- and Wald-type tests for $\tau = 0.5$, $\epsilon = 0.15$ and $T = 1000$.

In large samples (cf. Table 3) the tests behave qualitatively almost identical for all DGPs. They attain reasonable power even when the break size is still low. Except for breaks in the ARCH coefficient (DGP 3) all DGPs exhibit monotonic power functions. Nevertheless, they behave quantitatively different. In case of a break in the unconditional variance, DGP 1, the DP is slightly superior bearing in mind that the LS tests are slightly oversized. Within the LS test the expWald > expLM > supWald and supLM. For a break in the conditional variance (DGPs 2-4) the supLM performs worst. Concerning the double shift DGPs 5 and 6 the sup and mean tests are slightly inferior to the exp version of the tests. The DP test also performs quite well in this double-shift context. Regarding the smoothly inflating shifts in DGPs 7-9 the mean version of the tests seems to be superior. The sup version and the DP test (i.e. the tests allowing for a breakpoint estimation) perform rather similarly for smoothly inflating volatilities.
All in all the Wald tests are superior to the DP and the functionals of the LM tests, whereas the expWald performs best.

3.3 Non-monotonic power

In his seminal article Vogelsang [1999] discusses the issue of non-monotonic power of CUSUM, LM and Wald tests when testing for a mean shift in time series. The non-monotonicity is caused by the LRV estimation of the process. If the bandwidth is estimated via the data-dependent method of Andrews [1991], excessive lags are chosen in the AR(1) approximation as the AR coefficient is biased towards one causing the LRV to become very large.

Robust alternatives concerning the LRV estimation resulting in monotonic power functions have i.a. been proposed by Juhl and Xiao [2009], Kejriwal [2009] or Yang and Vogelsang [2011]. In terms of testing for breaks in volatility on the other hand, Xu [2013a] shows that the AR coefficient is no longer biased resulting in monotonic power functions for the tests. Concretely he argues that once the mean of the squared series is subject to a structural change, the same applies to the volatility of the squared series which prevents the long-run kurtosis estimator from selecting excessive lags.

In the present case $\Theta = \gamma_0 + 2\sum_{r=1}^{\infty}\gamma_r$ marks the LRV of η_t in all tests. Hence, in order for Θ to be finite the autocovariances γ_r have to be finite. That is for $r > 0$, $E(\eta_t^2) < \infty$ while for $r = 0$ additionally $E(\eta_t^4) < \infty$ has to be fulfilled $\forall\ t = (1, \ldots, T)$. If the latter condition is not fulfilled the tests suffer from non-monotonic power leading to the following theorem.

Theorem 1. *Let $\{\eta_t\}$ be a mean-zero $\alpha-$mixing stochastic process with bounded second moment $E(\eta_t^2) = \sigma^2 < \infty$. If $E(\eta_t^4) \to \infty$ we have that $CUSQ \to 0$, $LM \to 0$ and $Wald \to 0$ even for increasing volatility breaks.* □

Actually, m tends to zero as $E(\eta_t^4) \to \infty$ reducing the LRV estimator to $\hat{\gamma}_0$. Hence, decreasing power arises through the fact that the moment condition fails rather than through a bias in the estimation of the AR(1) coefficient for the bandwidth selection.

As a simple example consider the GARCH(1,1) process $\eta_t = h_t u_t$ with $h_t^2 = \omega + \alpha u_{t-1}^2 + \beta h_{t-1}^2$ and *iid* innovation $\{u\}$ with $E(u_t) = 0$ and $E(u_t^2) = 1$. Then $E(\eta_t^{2m})$ is only given under the condition that $\sum_{i=0}^{m} m!\alpha^i\beta^{m-i}E(u_t^{2i})/((m-i)!i!) < 1$, cf. He and Teräsvirta [1999]. As

$\mu_2 \equiv E(\eta_t^2) = \alpha + \beta$ and $\kappa \equiv E(\eta_t^4) = \beta^2 + 2\alpha\beta + \alpha^2 E(u^4)$, and noting that $\kappa < 1$ implies that $\mu_2 < 1$, $\Theta < \infty$ only if $\kappa < 1$.

Consider e.g. the case where $\alpha = \beta = 0.4$ with normally distributed errors. We then have $\mu_2 = 0.8$ and $\kappa = 0.96$. If however α switches to 0.5, $\kappa = 1.31$ and the upper considerations imply that the tests would have decreasing power for detecting the break in unconditional variance.

To underline these considerations we carry out some simple simulations seeking the varying coefficient GARCH(1,1) which has also been utilized by Hillebrand [2005] and Xu [2013a]. The process is given by $\eta_t = h_t u_t$ where $h_t = \omega + \alpha_t u_{t-1}^2 + \beta h_{t-1}^2$ and $u \overset{iid}{\sim} N(0,1)$. The ARCH coefficient is time dependent under the alternative switching from α_0 to α_1 a time $[\tau T]$ with $\tau \in [0,1]$, that is $\alpha_t = (\alpha_0 + (\alpha_1 - \alpha_0))\mathbb{1}_{\{t \geq [\tau T]\}}$. We test for a break in the unconditional variance process $\sigma_t^2 \equiv \eta_t^2$, i.e. $H_0 : \sigma_t^2 = \sigma^2$ vs. $H_1 : \sigma_t^2$ is not constant over t for the supLM test.[3]

We consider four DGPs based on different specifications of the GARCH parameter β under the Null: (i) DGP1: $\beta = 0$, (ii) DGP2: $\beta = 0.4$, (iii) DGP3: $\beta = 0.75$ and (iv) DGP4: $\beta = 0.75$ and $u \sim t(8)$, where $\alpha_0 = 0.1$ in all specifications. (i) describes the ARCH(1) specification being used by Deng and Perron [2008], (ii) has been considered in Xu [2013a] and Xu [2013b] while (iii) and (iv) are persistent versions of (ii). ω is specified such that the annualized volatility $\Sigma = \sqrt{250\omega/(1-\alpha-\beta)}$ equals 0.1 under the Null.

[3]Results for the supLM test are available upon request.

α_1	κ	Σ	$T = 500$		$T = 1000$		$T = 2000$	
			DP	LM	DP	LM	DP	LM
0.1	0.03	0.10	0.037	0.038	0.037	0.038	0.040	0.040
0.2	0.12	0.11	0.071	0.069	0.125	0.107	0.233	0.197
0.3	0.27	0.11	0.157	0.142	0.361	0.292	0.713	0.641
0.4	0.48	0.12	0.276	0.234	0.630	0.548	0.935	0.898
0.5	0.75	0.13	0.381	0.318	0.756	0.691	0.951	0.937
0.6	1.08	0.15	0.455	0.392	0.797	0.743	0.915	0.894
0.7	1.47	0.17	0.481	0.414	0.753	0.708	0.875	0.858
0.8	1.92	0.21	0.465	0.412	0.675	0.636	0.791	0.769
0.9	2.42	0.20	0.436	0.392	0.597	0.571	0.706	0.690
1.0	3.00	∞	0.409	0.386	0.558	0.540	0.626	0.616

Table 4: displays the power of the DP and LM test under volatility shifts with $h_t = \omega + (\alpha_0 + (\alpha_1 - \alpha_0))\mathbf{1}_{t \geq [\lambda T]} u_{t-1}^2$, $\lambda = 0.5$, $\omega = 3.6e-05$ and $\alpha_0 = 0.1$.

α_1	κ	Σ	$T = 500$		$T = 1000$		$T = 2000$	
			DP	LM	DP	LM	DP	LM
0.10	0.27	0.10	0.057	0.062	0.038	0.043	0.044	0.046
0.20	0.44	0.11	0.222	0.201	0.120	0.111	0.239	0.200
0.30	0.67	0.13	0.572	0.519	0.359	0.298	0.709	0.636
0.35	0.81	0.14	0.708	0.655	0.505	0.434	0.858	0.802
0.40	0.96	0.16	0.775	0.729	0.627	0.542	0.938	0.899
0.45	1.13	0.18	0.814	0.776	0.725	0.646	0.951	0.930
0.50	1.31	0.22	0.790	0.765	0.759	0.690	0.948	0.931
0.55	1.51	0.32	0.769	0.742	0.792	0.723	0.934	0.918
0.60	1.72	∞	0.781	0.778	0.785	0.735	0.918	0.903

Table 5: displays the power of the DP and LM test under volatility shifts with $h_t = \omega + (\alpha_0 + (\alpha_1 - \alpha_0))\mathbf{1}_{t \geq [\lambda T]} u_{t-1}^2 + 0.4 h_{t-1}$, $\lambda = 0.5$, $\omega = 2e-05$ and $\alpha_0 = 0.1$.

α_1	κ	Σ	$T = 500$		$T = 1000$		$T = 2000$	
			DP	LM	DP	LM	DP	LM
0.100	0.74	0.10	0.217	0.223	0.250	0.254	0.282	0.293
0.125	0.80	0.11	0.346	0.341	0.472	0.465	0.643	0.635
0.150	0.86	0.12	0.596	0.576	0.812	0.797	0.956	0.947
0.175	0.92	0.14	0.820	0.794	0.965	0.957	0.998	0.997
0.200	0.98	0.17	0.931	0.914	0.995	0.992	0.999	0.998
0.225	1.05	0.25	0.974	0.962	0.992	0.991	0.997	0.994
0.250	1.13	∞	0.983	0.981	0.994	0.993	0.996	0.995

Table 6: displays the power of the DP and LM test under volatility shifts with $h_t = \omega + (\alpha_0 + (\alpha_1 - \alpha_0))\mathbf{1}_{t \geq [\lambda T]} u_{t-1}^2 + 0.75 h_{t-1}$, $\lambda = 0.5$, $\omega = 6e-06$ and $\alpha_0 = 0.1$.

α_1	κ	Σ	$T = 500$		$T = 1000$		$T = 2000$	
			DP	LM	DP	LM	DP	LM
0.100	0.76	0.10	0.188	0.204	0.229	0.239	0.257	0.266
0.125	0.82	0.11	0.272	0.276	0.386	0.383	0.512	0.497
0.150	0.89	0.12	0.456	0.450	0.654	0.630	0.840	0.821
0.175	0.96	0.14	0.662	0.628	0.863	0.842	0.978	0.970
0.200	1.04	0.17	0.820	0.795	0.954	0.944	0.992	0.990
0.225	1.13	0.25	0.909	0.885	0.978	0.972	0.993	0.990
0.250	1.22	∞	0.897	0.881	0.967	0.963	0.979	0.975

Table 7: displays the power of the DP and LM test under volatility shifts with $h_t = \omega + (\alpha_0 + (\alpha_1 - \alpha_0))\mathbf{1}_{t \geq [\lambda T]} u_{t-1}^2 + 0.75 h_{t-1}$, $\lambda = 0.5$, $\omega = 6e-06$, $u \sim t(8)$ and $\alpha_0 = 0.1$.

Under the alternative $\alpha_1 \in [0.1, 1-\beta]$, specifying an IGARCH model at the upper limit of the interval. We consider three different sample sizes of $T = (500, 1000, 2000)$ with breakpoint specifications of $\tau = (0.5, 0.9)$[4]. The number of replications is $M = 5000$. Tables 4-7 return size and power results for the four DGPs.

For the ARCH(1) in Table 4 one can clearly see for both tests that the power becomes larger with an increasing ARCH coefficient up to a value of about 0.6 (0.7) in large (small) samples only to decrease if the break becomes bigger. In fact the power increases with increasing κ given that κ is smaller than one and decreases once the condition is no longer fulfilled, i.e. once the LRV of the squared process is no longer finite. This corresponds to the upper considerations that the LRV becomes infinite for $\kappa > 1$ leading to a power drop in both tests. Furthermore the power seems to be the lower the higher the value of κ gets - regardless of the size of the switch in annualized volatility. Concerning the sample size the power drop occurs earlier and is bigger for large T. As obviously more observations are drawn in large samples the probability of drawing a large outlier causing the LRV to converge to infinity is also higher for large T which results in the diverse behavior of the power concerning T.

Tables 5 to 7 support these findings although the power drop is not as big as for DGP1. This may be due to the fact that κ cannot reach such large values as in DGP1. E.g. in DGP3 $\max \kappa = 1.13$ which in finite samples does not seem to imply a convergence of Θ to infinity fast enough to lead to a drop in power. We can however at the least observe a "stagnation" of power in all DGPs once $\kappa > 1$ in spite of increasing annualized volatility.

3.4 Density estimation of break points

In order to assess how correctly the real breakpoint τ is estimated in the employed testing procedures we want to plot the density of the true break point estimators. Hence, only the testing procedures whose test statistic is based upon a supremum are being considered here, namely the DP-, supLM and supWald-test. Since DGPs of process type III lack a distinct break point and double shifting processes do not qualify well (DGPs 5-6) only DGPs 2-4 are under consideration.[5]

For $T = 200$ and $\tau = 0.5$ we conduct the CUSQ-test $M = 5000$ times and compare the test statistic to the 5% critical value of 1.358. For our purpose we assume the maximum break specification, such that $\Sigma = 2$, meaning that c.p. the processes switch as follows:

$$\begin{array}{llcl} \text{DGP 2:} & \mu_0 = 2 \cdot 10^{-5} & \rightarrow & \mu_1 = 8 \cdot 10^{-5} \\ \text{DGP 3:} & \alpha_0 = 0.1 & \rightarrow & \alpha_1 = 0.48 \\ \text{DGP 4:} & \beta_0 = 0.4 & \rightarrow & \beta_1 = 0.78. \end{array} \quad (5)$$

Within the sample the break point value implied by the maximum test statistic that exceeds the critical value is chosen as our estimated break point. For all three τ we plot the estimated τ^* and obtain the nonparametric density estimator of the break point estimators $f(\tau^*)$, whereby the dotted lines depict the mode of $f(\tau^*)$. Figure 3 plots the

[4]We report only the results for $\tau = 0.5$ here. Results for $\tau = 0.9$ are available upon request.
[5]Results for DGP 1 and DGPs 5-6 are available upon request.

density for $T = 200$.

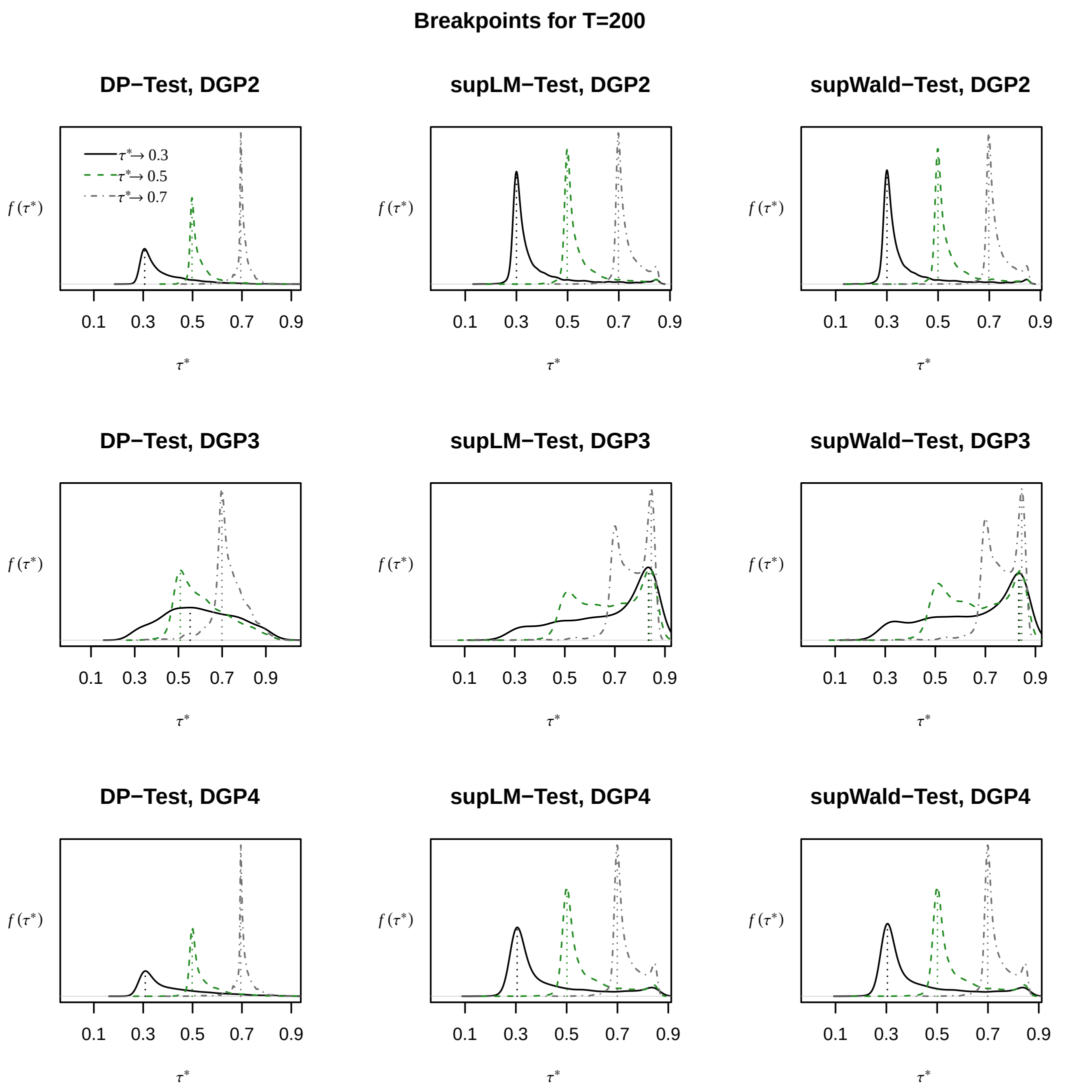

Figure 3: displays the density of the breakpoints in case of DGP 2-4 for $T = 200$, $\Sigma = 2$ and all τ.

At first sight the DGPs behave qualitatively analogical over all three tests, which is why we differentiate between the DGPs instead of the tests in the following discussion. We also yield qualitatively similar results for the different sample sizes not shown here, but can be found in the appendix. For DGP 2 we obtain leptokurtic densities with a peak centered at the true τ, indicated by the dotted lines. The densities especially of the least-squares tests are positively skewed. The results for DGP 4 show likewise patterns as in case of DGP 2 but not as good as for DGP 2. The densities are less peaked around the insinuated τ and yet more positively skewed. They even show a tendency for a multimodal distribution, which is due to the truncation at the upper bound with $\epsilon = 0.15$ and hence, $\tau = 0.85$. Notably, the least-squares tests perform very alike for all DGPs. However, in case of DGP 3 we have to discuss the results for all τ separately. For $\tau = 0.3$ the density is neither leptokurtic nor centered around the assumed τ and altogether performs worst particularly in regard of the DP test. It is noticeable that in

	τ	DP 0.3	DP 0.5	DP 0.7	*sup*LM 0.3	*sup*LM 0.5	*sup*LM 0.7	*sup*Wald 0.3	*sup*Wald 0.5	*sup*Wald 0.7
DGP 2					$T = 200$					
	$x_{0.5}$	0.340	0.510	0.700	0.320	0.515	0.715	0.320	0.515	0.715
	sd	0.096	0.048	0.033	0.125	0.081	0.047	0.125	0.081	0.048
					$T = 500$					
	$x_{0.5}$	0.318	0.504	0.700	0.306	0.506	0.704	0.306	0.506	0.704
	sd	0.058	0.022	0.014	0.054	0.036	0.028	0.054	0.036	0.028
					$T = 1000$					
	$x_{0.5}$	0.309	0.504	0.700	0.302	0.502	0.702	0.392	0.560	0.737
	sd	0.032	0.022	0.008	0.016	0.016	0.015	0.192	0.118	0.054
DGP 3					$T = 200$					
	$x_{0.5}$	0.590	0.590	0.720	0.745	0.695	0.775	0.710	0.670	0.770
	sd	0.162	0.115	0.077	0.174	0.130	0.072	0.181	0.132	0.074
					$T = 500$					
	$x_{0.5}$	0.466	0.550	0.714	0.516	0.594	0.750	0.500	0.592	0.750
	sd	0.167	0.098	0.059	0.203	0.127	0.059	0.202	0.126	0.059
					$T = 1000$					
	$x_{0.5}$	0.404	0.535	0.710	0.390	0.560	0.737	0.392	0.560	0.737
	sd	0.154	0.085	0.042	0.192	0.118	0.054	0.192	0.118	0.054
DGP 4					$T = 200$					
	$x_{0.5}$	0.360	0.510	0.700	0.335	0.525	0.715	0.335	0.520	0.715
	sd	0.131	0.066	0.041	0.170	0.103	0.054	0.169	0.103	0.054
					$T = 500$					
	$x_{0.5}$	0.334	0.506	0.700	0.312	0.510	0.708	0.312	0.510	0.708
	sd	0.088	0.038	0.020	0.113	0.073	0.041	0.113	0.073	0.041
					$T = 1000$					
	$x_{0.5}$	0.317	0.503	0.700	0.304	0.504	0.704	0.304	0.504	0.704
	sd	0.061	0.023	0.010	0.060	0.045	0.027	0.060	0.045	0.027

Table 8: reports the median an standard deviation for the breakpoints of all sample sizes over all tests for DGPs 2-4.

case of $\tau = 0.5$ and 0.7 the density is even bimodal for the least squares tests and that the peak centers around $\tau = 0.85$ for all three τ, where the sample is truncated. The reason is that DGP 3 exhibits outliers causing the jump in the variance ($\alpha_0 = 0.1 \rightarrow \alpha_1 = 0.48$) and suffers from an infinite kurtosis as already pointed out in the preceding section, c.f. 3.3.

In large samples, $T = 1{,}000$, the density peaks around the insinuated τ for all τ^* in case of DGP 2 and DGP 4. The least squares test obtain slightly more leptokurtic densities compared to the DP test. While the positive skewness of the least squares tests is identical for all τ and nearly negligible, the positive skewness of the DP test declines in τ. In case of DGP 3 the DP break point estimator performs slightly better, since the densities for all τ are unimodal and again, the positive skewness declines in τ. Although the least squares tests peak around the true τ, in large samples they nevertheless have a second mode at the truncation point. All in all is the DP break point estimator slightly better in case of DGP 3 in large samples, but yields little less good results than the least squares counterparts in case of DGP 2 and DGP 4.

To confirm the findings from Figure 3 Table 8 describes the median and the standard

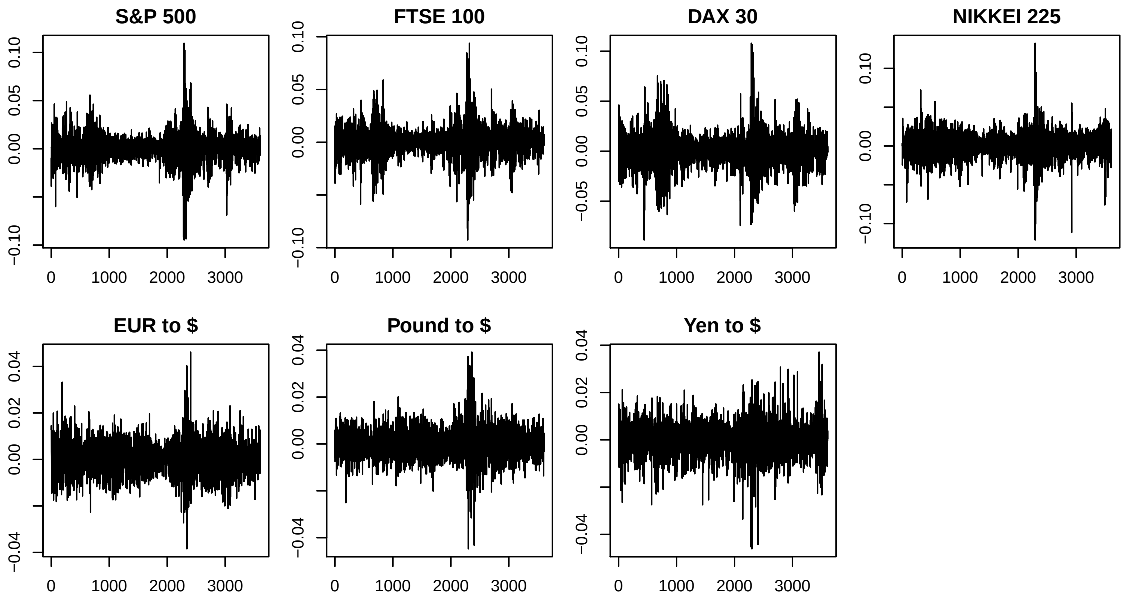

Figure 4: plots the returns of the 7 time series.

deviation of the break point estimators according to the previous graphic. It is striking that the standard deviation for $\tau = 0.3$ is always the highest for all three tests. The earlier the break happens the higher is the dispersion of the break point estimator. Because of an early break the probability for a maximum peak within the remaining sample period is high, since the sample fluctuations increase significantly after the break. Hence, it is probable that a higher peak than the break point occurs and is chosen as the supremum of the test statistic. On the other hand is it more unlikely to find a higher peak than the break point, when the break happens at $\tau = 0.5$ or $\tau = 0.7$.

4 Empirical Analysis

As an empirical example we consider 7 financial time series, namely the returns of 4 stock market indices (S&P 500, FTSE 100, DAX 30, Nikkei 225) and 3 exchange rates (Euro, Pound and Yen) to the Dollar. We have daily data taken from *Datastream* from 01/01/2000 until 10/30/2013 yielding $T = 3{,}608$ observations for each variable. Figure 4 returns a plot of the data.

In order to get a first idea of the behavior of the tests we determine the volatility break points for each series conducting the DP, the supLM and the supWald test. The breakpoint is estimated via the iterated cumulative sums of squares (ICSS) algorithm of Inclan and Tiao [1994] where $\alpha = 0.05$ throughout the analysis.

Defining $\mathcal{J}_i(\tau)$ as the value of the statistic of test i at $\tau \in [0,1]$, the breakpoint estimator in the single break case is simply defined as the point where the maximum of the respective test statistic, conditional on rejecting the Null, is reached. Hence, $\tau^* := \arg\max_{1\leq\tau\leq T} \mathcal{J}_i(\tau) \,|\, \mathcal{J}_i(\tau^*) > Q_i^\alpha$ where Q_i^α marks the critical value of test i and level α.

In the multiple break case this procedure is carried out iteratively. Starting with one

	# Breaks	Break Date 1	2	3	4	5	6	7
					S&P 500			
DP	4	06/17/02	05/20/03	09/04/08	12/21/11			
*sup*LM	5	07/05/02	04/03/03	09/04/08	06/12/09	12/21/11		
*sup*Wald	5	07/05/02	04/03/03	09/04/08	06/12/09	12/21/11		
					FTSE 100			
DP	7	06/12/02	04/17/03	07/23/07	04/06/09	12/15/11	08/06/12	05/27/13
*sup*LM	6	08/06/01	06/14/02	04/18/03	07/23/07	04/06/09	08/06/12	
*sup*Wald	6	08/06/01	06/13/02	04/18/03	07/23/07	04/06/09	08/06/12	
					DAX 30			
DP	7	08/30/01	06/14/02	06/17/03	01/21/08	07/16/09	08/01/11	08/06/12
*sup*LM	6	08/30/01	06/14/02	05/20/03	01/15/08	04/03/09	08/06/12	
*sup*Wald	6	08/30/01	06/14/02	05/20/03	01/15/08	04/03/09	08/06/12	
					NIKKEI 225			
DP	3	12/18/03	01/04/08	05/20/09				
*sup*LM	3	12/18/03	01/04/08	12/16/08				
*sup*Wald	3	12/18/03	01/04/08	12/16/08				
					$/€			
DP	4	09/26/01	08/16/04	08/11/08	11/16/11			
*sup*LM	4	04/23/01	08/11/08	05/25/09	03/12/12			
*sup*Wald	4	04/23/01	08/11/08	05/25/09	03/12/12			
					Ł/$			
DP	4	06/22/01	01/05/04	08/08/08	11/15/11			
*sup*LM	5	04/20/01	01/05/04	08/08/08	06/10/09	11/24/11		
*sup*Wald	5	04/20/01	01/05/04	08/08/08	06/10/09	11/24/11		
					¥/$			
DP	3	08/07/07	08/17/09	05/02/11				
*sup*LM	4	06/07/06	08/07/07	04/01/09	02/21/13			
*sup*Wald	4	06/07/06	08/07/07	04/01/09	02/21/13			

Table 9: returns the estimated break dates for the 3 tests and 7 series. The dates were estimated conducting the ICSS algorithm of Inclan and Tiao [1994] with $\alpha = 0.05$ and $\varepsilon = 0.15$.

breakpoint the sample is divided around this very point and the test is implemented within both subsamples. If further breakpoints are detected this procedure is repeated until the test cannot reject any more. Additionally a minimum segment size h should be specified in advance. We set $h = 200$, i.e. we allow breaks to occur every 10 months at most. Furthermore breaks are allowed to occur in the interval $\tau \in [\varepsilon, 1-\varepsilon]$ where ε is again specified as 0.15. The results are given in Table 9 displaying the number of estimated breaks and the corresponding break dates for the three tests over the seven series.

As one can clearly see the supLM and the supWald test yield (with very few exceptions) the same results. Not only are the number of breaks identical for each series, also the estimated break dates do almost not differ between the tests. This is, of course, not very surprising regarding the similarity of the test statistics.[6]

The DP test however leads to different results in some situations. For the S&P 500 and two exchange rates fewer breaks are found. On the other hand the DP test detects more breaks for the DAX 30 and the Nikkei 225. If a break is found within some time period

[6]We also reduced the window size to $h = 100$. The supLM and supWald tests still did not differ in the number of detected breaks. The breakpoint estimation however varied much stronger.

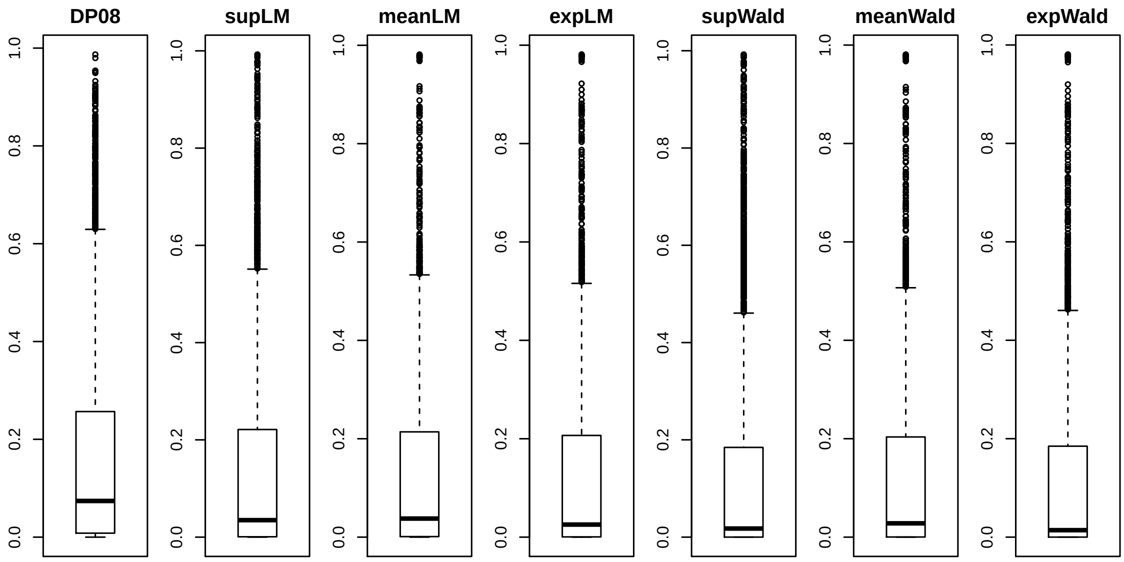

Figure 5: displays the boxplots for the p-values of the respective test on constant volatility for the S&P 500 with $\varepsilon = 0.15$.

by all three tests the estimated break dates do not differ much over the tests. There also seems to be a tendency of (on average) fewer breaks in volatility in the currencies than in the equity series. Hence, the number and the timing of the break does indeed differ considerably between the tests.

In order to get a more detailed look into the behavior of the tests, we consider rolling window estimations for each test over the seven series. Hence, we determine the test statistics and the corresponding p-value for each window for each test and series and are thus able to examine and compare the distribution of the p-values over the different tests.

Concretely we test for a single break within a window of size $h = 200$ with $\varepsilon = 0.15$. That is, for each test and series we derive 3,408 p-values and are thus able to compare the distribution of the latter between the different tests applied to real data. Figures 5 and 6 return boxplots for the p-values of each volatility break test for the S&P 500 and the $/€ series.[7] Table 10 displays the rejection frequencies over the series and tests. That is, specifying $\alpha = 0.05$ the table returns for how many per cent of the 3,408 statistics the hypothesis of constant volatility has been rejected.

The median of the p-value of the DP test is clearly higher for each series compared to the other tests. Hence, the LM and Wald tests tend to reject more often than the DP test in this environment. Furthermore the p-values of the DP test tend to vary more in between its lower and upper quartile whereas the variation in the tails seems to be higher for the LM and Wald tests. Hence, we observe more variation within the "core" of the distribution for the DP test whilst a larger amount of outliers occurs for the LM and Wald tests.

[7]Results for the remaining series can be found in the appendix.

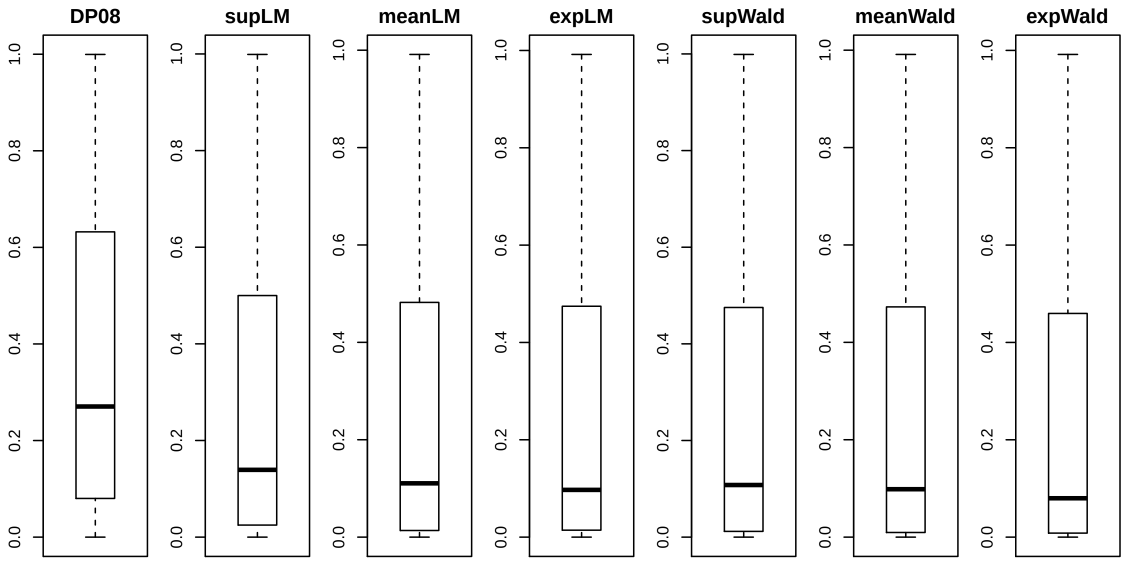

Figure 6: displays the boxplots for the p-values of the respective test on constant volatility for the \$/€ exchange rate with $\varepsilon = 0.15$.

Furthermore it seems noticeable that for the currency series the p-values of the tests are substantially larger compared to the equity series. This also corresponds to the upper findings that on average there are fewer breaks in the currency series than in the equity series. By recalling Figure 4 we note that volatility shifts seem to be smoother in the currency data than for the equity series, since for the latter a distinctive clustering effect in volatility can be observed. Additionally the simulation study showed that for smooth transitions the power of detecting a volatility shift is rather low which may be a possible explanation for these effect.

Hence, we can conclude that there may indeed be some severe differences between the tests when it comes to break detection and estimation even in real data examples. First the DP test tends to reject the hypothesis of constant volatility less often than LM and Wald tests for the existing data. Regarding the €/\$ series the range even amounts to 26%. Second all tests seem to exhibit less power for detecting a volatility shift when the break is rather smooth than abrupt. This may be exemplified by the finding that in the currency series much fewer volatility shifts are found than in the equity series.

5 Conclusion

In this paper we analyze volatility break tests by conducting a simulation analysis as well as empirical examples using equity and exchange rate data. Concerning the simulations we find that for some DGPs the difference over the tests is rather high whereas for other DGPs it does not seem to play an important role which test is utilized.

In small samples the expWald test exhibits the highest power. However, it is slightly oversized. In large samples the difference is not as distinct. Still, for double shifting DGPs the DP test seems to be superior to the other tests. In case of DGP 3 non-

	Test							
	DP	*sup*LM	*mean*LM	*exp*LM	*sup*Wald	*mean*Wald	*exp*Wald	Range
S&P 500	0.44	0.54	0.54	0.59	0.61	0.57	0.63	0.19
FTSE 100	0.41	0.56	0.58	0.59	0.61	0.61	0.63	0.22
DAX 30	0.51	0.64	0.64	0.68	0.71	0.67	0.73	0.22
NIKKEI 225	0.40	0.46	0.44	0.50	0.52	0.47	0.53	0.13
€/$	0.18	0.33	0.40	0.40	0.39	0.42	0.44	0.26
Ł/$	0.22	0.36	0.39	0.40	0.41	0.41	0.44	0.22
¥/$	0.17	0.24	0.26	0.28	0.31	0.29	0.33	0.16

Table 10: returns the rejection frequencies concerning the respective test on constant volatility with $\alpha = 0.05$ and $\varepsilon = 0.15$. The frequencies are calculated on the basis of 3,408 rolling window estimations for each test and series.

monotonic power functions in large samples are observed for all tests. In this process the ARCH coefficient switches in such a way that for large breaks the kurtosis of the process is no longer finite. As a consistent estimation of the long-run variance of the squared process depends on the finite assumption the power eventually drops once this assumption is no longer fulfilled.

Regarding the empirical example we find that less breaks are found in the exchange rate data than in the equity data. This may be caused by the fact that we rather observe a smooth behavior of volatility and not a distinct clustering behavior in the exchange rate data. Another reason could be that the exchange rate data may have more outliers instead of clustering behavior. As the simulations show in these situations the tests perform rather poorly. Additionally we perform rolling window estimations in order to compare the p-values of the tests over a broad range of window estimations. Hereby the results of the simulation study are confirmed. Further we find a substantially lower amount of breaks in the currency data.

All in all perform the least squares tests in most of the situations fairly similar which is due to the fact, that the test statistics are very akin. But in regard of the power we can state that the Wald test can be slightly superior. Since it becomes more difficult to differentiate between break and outlier when the volatility shift occurs rather smoothly than abrupt, all of the tests seem to fail to appropriately detect breaks, cf. 3.2 and 4. Nevertheless, were we able to derive slightly better results for the DP test especially when the least squares tests miscarry in case of DGP 3, cf. 3.4. Hence, there is a point in choosing a particular test, depending on the tested situation.

To conclude it could be of interest for future work to be able to state if the volatility break happens in the conditional or unconditional variance or further to tell if the break occurs in higher moments or yet in distribution.

References

C. Amado and T. Teräsvirta. Modelling volatility by variance decomposition. *Journal of Econometrics*, 175:142–153, 2013.

E. Andreou and E. Ghysels. Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics*, 17:579–600, 2002.

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

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

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

R.L. Brown, J. Durbin, and J.M. Evans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 149–192, 1975.

G. Cavaliere and A. Taylor. Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics*, 140:919–947, 2007.

A. Deng and P. Perron. The limiting distribution of the cusum of squares test under general mixing conditions. *Econometric Theory*, 24:809–822, 2008.

F.X. Diebold. Modeling the perisistence of conditional variances: a comment. *Econometric Reviews*, 5:51–56, 1986.

R. Denise D.O. van Dijk and M. Sensier. Testing for causality in variance in the presence of breaks. *Economics Letters*, 89(2):193–199, 2005.

R.F. Engle and J. Rangel. The apline GARCH model model for low frequency volatility and its global macroeconomic fundamentals. *Review of Financial Studies*, 21:1187–1222, 2008.

R.F. Engle, E. Ghysels, and B. Sohn. Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics*, 95(3):776–797, 2013.

C. He and T. Teräsvirta. Properties of moments of a family of GARCH processes. *Journal of Econometrics*, 92:173–192, 1999.

E. Hillebrand. Neglecting parameter changes in GARCH models. *Journal of Econometrics*, 129:121–138, 2005.

C. Inclan and G.C. Tiao. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89(427):913–923, 1994.

T. Juhl and Z. Xiao. Tests for changing mean with monotonic power. *Journal of Econometrics*, 148:14–24, 2009.

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

C.G. Lamoureux and W.D. Lastrapes. Persistence in variance structural change and the GARCH model. *Journal of Business and Economic Statsitics*, 8:225–234, 1990.

D.E. Rapach and J.K. Strauss. Structural breaks and garch models of exchange rate volatility. *Journal of Applied Econometrics*, 23(1):65–90, 2008.

M. Sensier and D. van Dijk. Testing for volatility changes in U.S. macroeconomic time series. *The Review of Economics and Statistics*, 86(3):833–839, 2004.

T.J. Vogelsang. Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics*, 88:283–299, 1999.

K.-L. Xu. Bootstrapping autoregression under nonstationary volatility. *Econometrics Journal*, 11:1–26, 2008.

K.-L. Xu. Power monotonicity in detecting volatility level changes. *Economics Letters*, 121:64–69, 2013a.

K.-L. Xu. Powerful tests for structural changes in volatility. *Journal of Econometrics*, 173:126–142, 2013b.

J Yang and T.J. Vogelsang. Fixed-b analysis of LM-type tests for a shift in mean. *The Econometrics Journal*, 14:438–456, 2011.

6 Appendix

A Figures

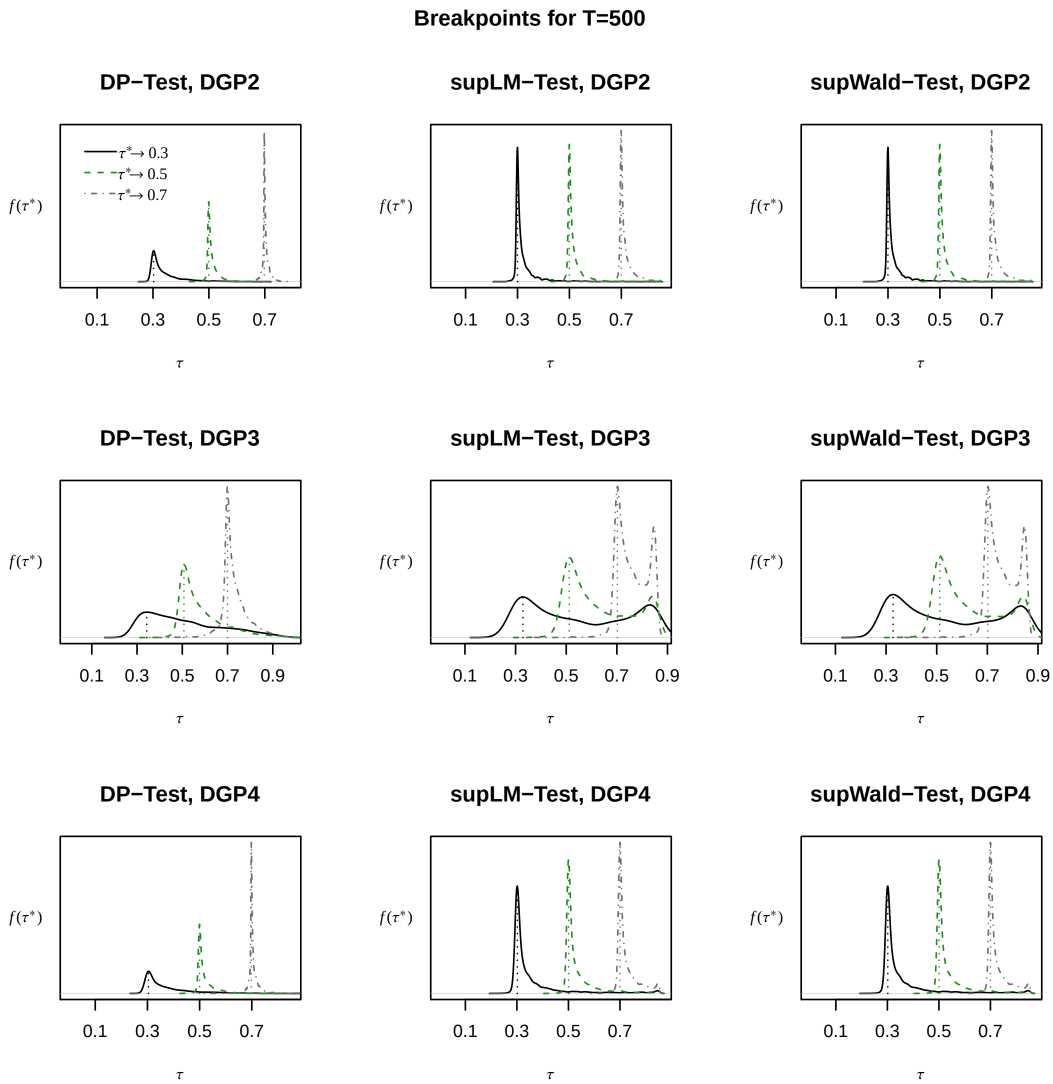

Figure 7: displays the density of the breakpoints in case of DGP 2-4 for $T = 500$, $\Sigma = 2$ and all τ.

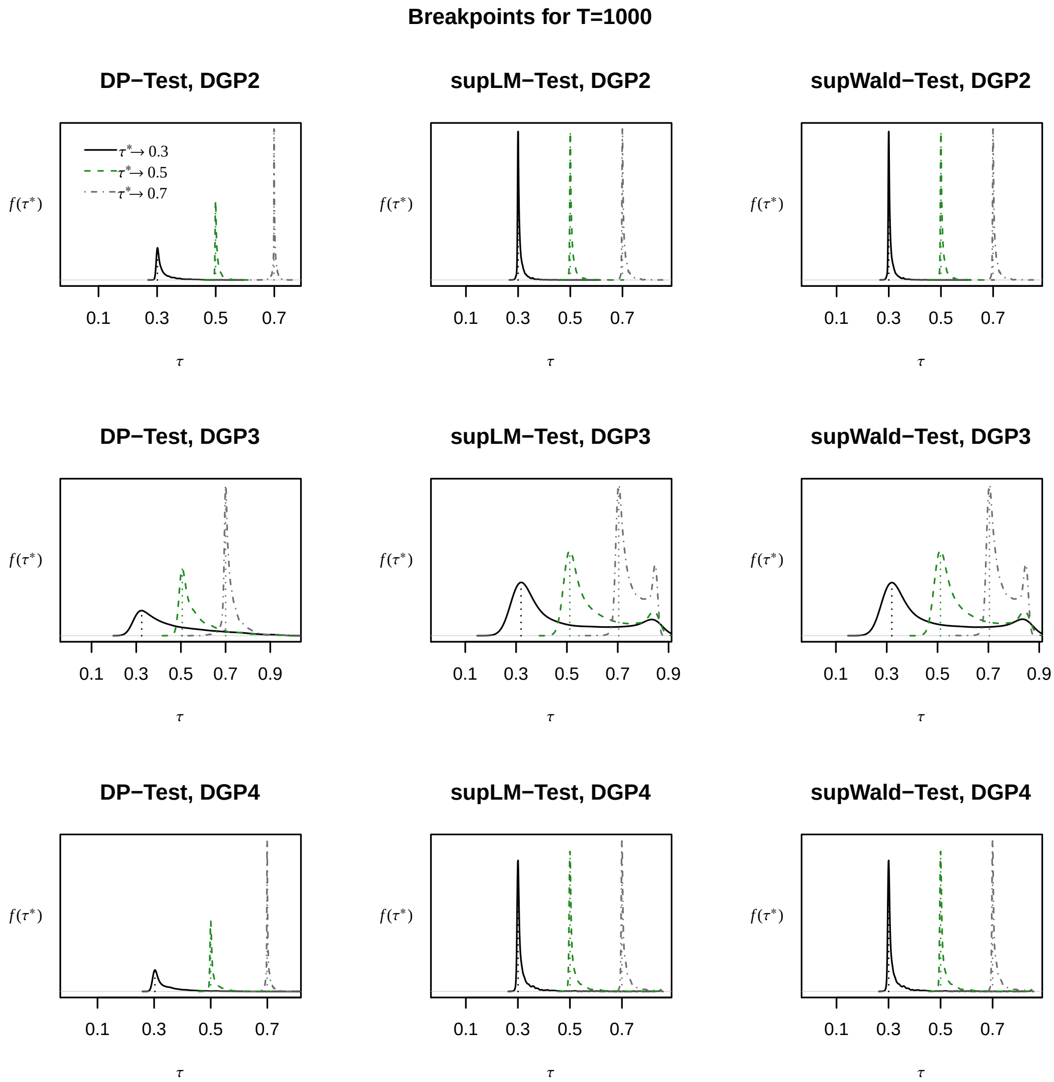

Figure 8: displays the density of the breakpoints in case of DGP 2-4 for $T = 1000$, $\Sigma = 2$ and all τ.

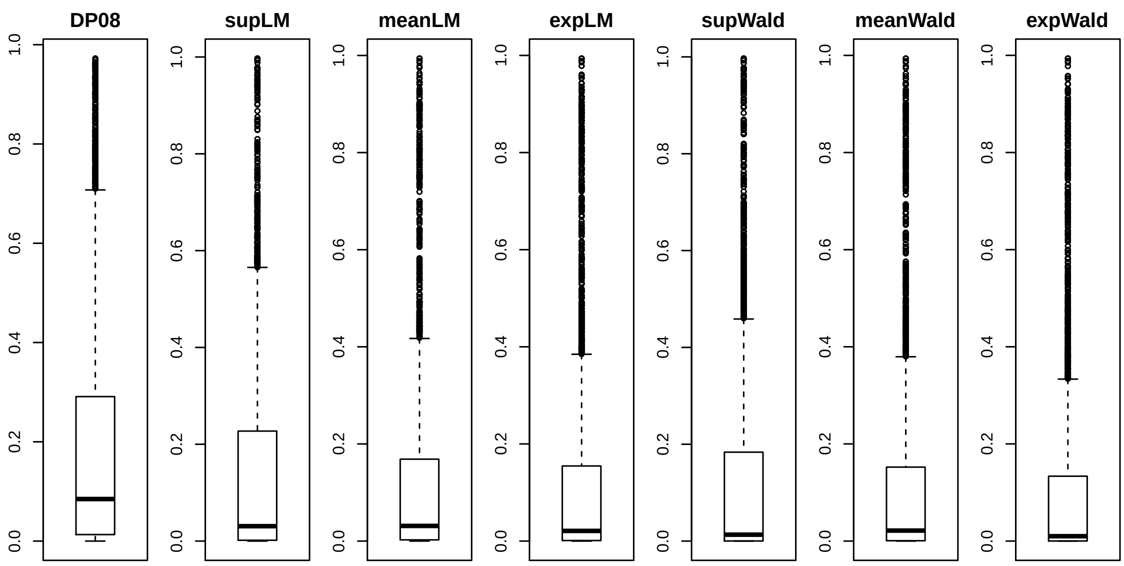

Figure 9: displays the boxplots for the p-values of the respective test on constant volatility for the FTSE 100 with $\varepsilon = 0.15$.

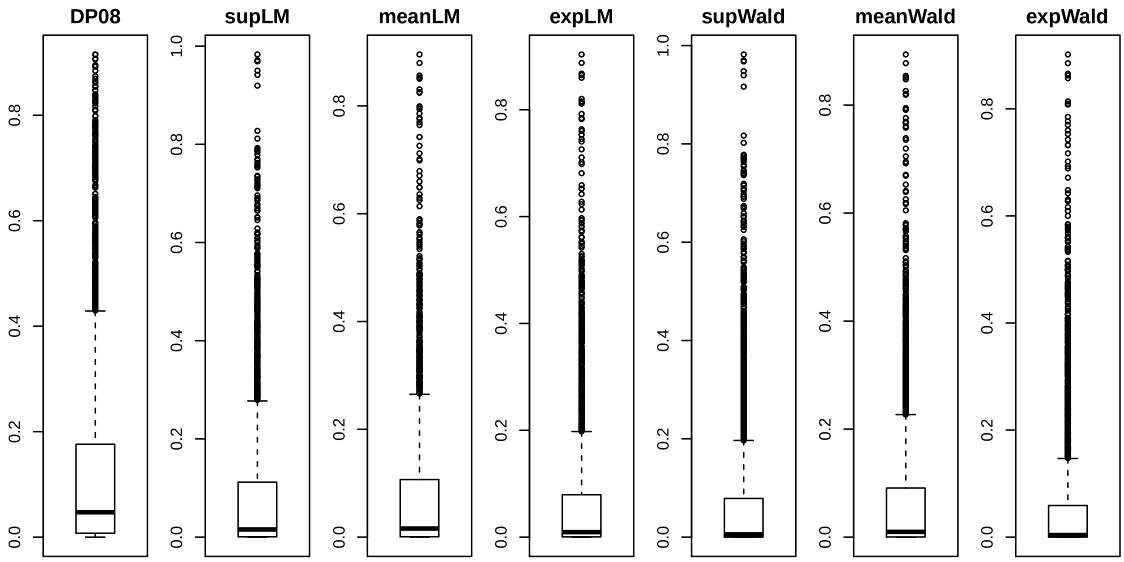

Figure 10: displays the boxplots for the p-values of the respective test on constant volatility for the DAX 30 with $\varepsilon = 0.15$.

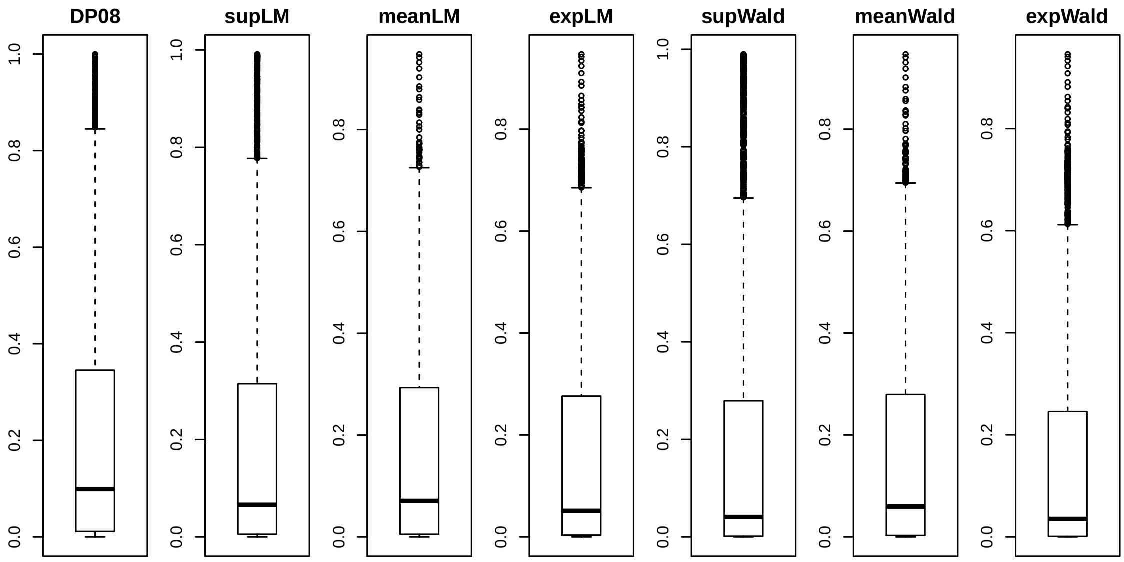

Figure 11: displays the boxplots for the p-values of the respective test on constant volatility for the Nikkei 225 with $\varepsilon = 0.15$.

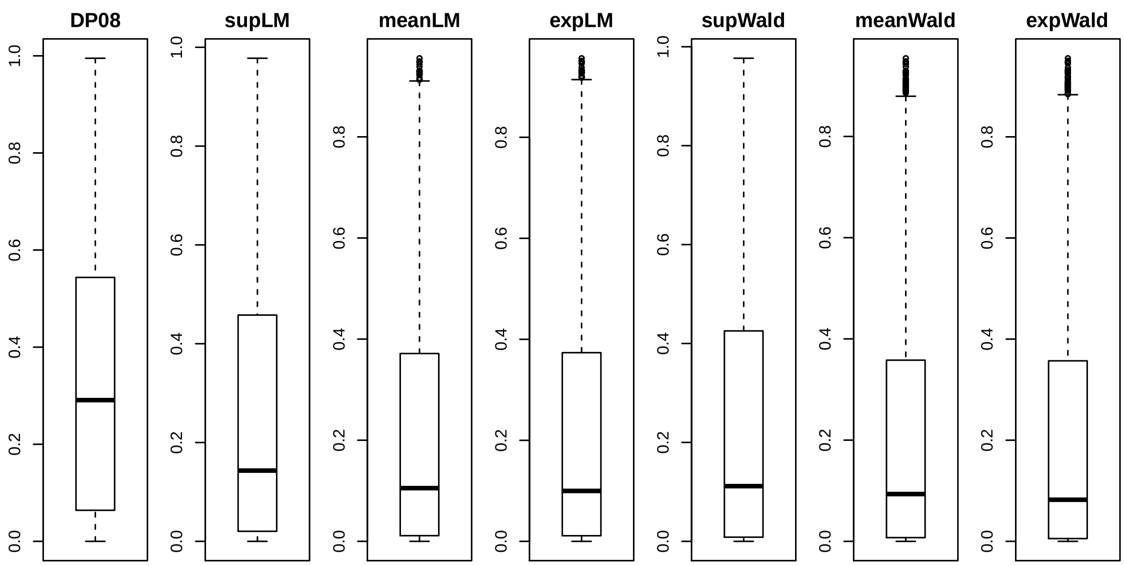

Figure 12: displays the boxplots for the p-values of the respective test on constant volatility for the Ł/$ exchange rate with $\varepsilon = 0.15$.

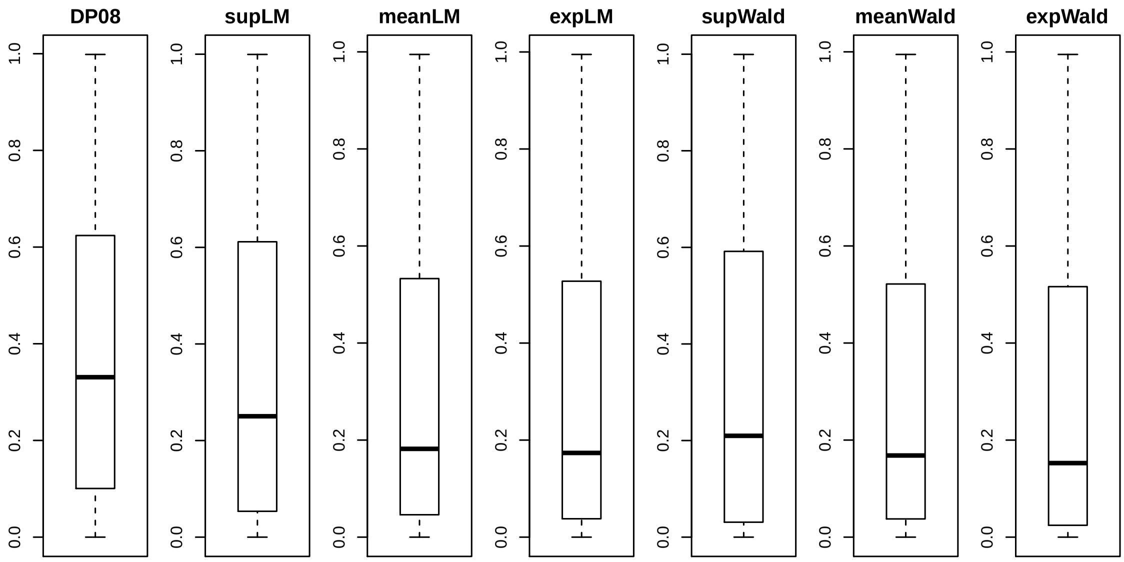

Figure 13: displays the boxplots for the p-values of the respective test on constant volatility for the ¥/$ exchange rate with $\varepsilon = 0.15$.

B Tables

	DGP 1							DGP 2							DGP 3						
		LM			Wald				LM			Wald				LM			Wald		
Σ	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp
1	6.06	6.54	7.20	7.60	8.30	8.20	8.30	5.50	5.50	6.12	6.42	6.40	6.00	7.40	5.36	5.50	6.12	6.42	6.40	6.00	7.40
1.1	20.07	17.28	22.90	21.50	20.30	24.60	23.30	20.10	17.38	22.50	21.70	22.30	25.20	25.90	17.44	16.16	19.80	19.08	21.00	22.20	23.20
1.2	58.12	50.62	59.24	57.76	54.40	60.00	60.30	58.28	50.96	59.36	58.22	54.70	60.00	59.30	38.86	34.76	41.46	40.26	38.50	42.20	42.70
1.3	88.04	83.68	87.70	88.00	88.40	89.90	90.60	88.26	83.54	88.00	88.04	86.00	89.10	89.40	58.28	52.42	61.10	60.56	58.90	65.00	64.60
1.5	99.78	99.58	99.80	99.78	99.70	99.80	99.80	99.82	99.50	99.78	99.82	99.60	99.80	99.80	74.96	69.54	78.10	77.70	74.60	81.10	81.20
2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	78.96	75.96	82.54	82.76	81.40	84.10	84.80

	DGP 4							DGP 5							DGP 6						
		LM			Wald				LM			Wald				LM			Wald		
Σ	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp
1	5.36	5.50	6.12	6.42	6.40	6.00	7.40	6.06	6.54	7.20	7.60	8.30	8.20	8.30	4.52	5.72	5.84	6.36	7.30	7.00	7.80
1.1	20.90	18.20	22.94	22.48	23.00	18.30	26.70	6.38	5.80	6.48	7.12	7.40	6.90	8.50	3.68	4.30	5.22	5.00	5.60	5.10	5.90
1.2	58.06	51.56	59.04	58.40	54.60	59.80	59.60	13.46	9.98	10.56	12.70	11.90	11.70	13.90	5.00	4.48	5.68	5.68	4.60	6.00	6.00
1.3	86.96	82.50	86.52	86.86	85.50	87.80	88.10	22.48	16.44	17.60	20.60	21.00	20.80	23.90	8.58	5.84	7.72	8.60	6.90	8.60	10.10
1.5	99.52	99.14	99.46	99.48	99.40	99.50	99.50	50.60	39.16	46.24	47.78	46.90	50.40	53.70	19.76	13.08	17.74	18.42	15.30	19.20	20.40
2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.08	87.46	95.74	94.84	89.20	96.20	95.90	54.68	40.98	56.30	54.02	46.90	59.00	58.20

	DGP 7							DGP 8							DGP 9						
		LM			Wald				LM			Wald				LM			Wald		
Σ	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp	DP	sup	mean	exp	sup	mean	exp
1	5.70	5.70	6.38	6.62	7.30	7.80	8.00	5.70	5.70	6.38	6.62	7.30	7.80	8.00	5.70	5.70	6.38	6.62	7.30	7.80	8.00
1.1	11.14	11.00	13.42	13.26	11.80	13.60	14.60	8.76	8.28	10.58	10.16	8.60	10.90	10.30	11.70	12.16	14.26	14.16	13.80	14.70	14.70
1.2	26.00	25.50	31.14	30.30	30.00	32.90	33.30	17.38	16.58	21.58	20.50	21.40	22.60	24.00	28.62	29.16	33.58	33.44	33.70	35.00	36.10
1.3	44.88	43.46	53.52	50.74	46.80	54.10	51.90	28.70	27.24	34.92	33.12	29.70	35.50	35.90	51.06	51.12	58.68	57.10	53.30	59.30	58.30
1.5	78.44	76.70	86.02	83.90	80.80	87.00	85.80	52.90	52.02	62.58	60.10	58.50	65.60	63.70	85.22	84.82	90.72	89.44	87.80	90.70	90.90
2	99.62	99.34	99.92	99.88	99.70	100.0	100.0	90.38	89.68	95.36	94.64	91.10	95.60	95.40	99.90	99.92	100.0	100.0	100.0	100.0	100.0

Table 11: reports power results for nine DGPs according to the DP, LM- and Wald-type tests for $\tau = 0.5$, $\epsilon = 0.15$ and $T = 200$.