Essays on Testing for Nonlinearity in Time Series -

Issues in Nonlinear Cointegration, Structural Breaks and Changes in Persistence

Von der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktorin der Wirtschaftswissenschaften
- Doctor rerum politicarum -

genehmigte Dissertation
von

Claudia Grote, M.Sc.
geboren am 08. September 1986 in Hannover

2020

Referent: Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover

Koreferent: Prof. em. Dr. Olaf Hübler, Leibniz Universität Hannover

Tag der Promotion: 04. März 2020

Danksagung

Diese Arbeit widme ich meinem Sohn Alexander.

Ich sage Dank in erster Linie meiner Mama, die stets eine gute Ausbildung sowie einen akademischen Abschluss gefördert und mit allen Mitteln unterstüzt hat. Meinem Doktorvater Philipp Sibbertsen bin ich zu herzlichem Dank verpflichtet, dass er bis zum Ende Vertrauen in mich hatte und mein Promotionsvorhaben befürwortet hat. Ganz besonderer Dank gebührt meinem engsten Vertrauten Johannes Rohde, ohne dessen Motivation ich wahrscheinlich niemals abgegeben hätte.
Ich danke außerdem allen, die an mich geglaubt haben und bin dankbar für die einzigartige Zeit am Institut, die unvergessen bleiben wird.

Claudia Grote
Hannover, März 2020

Zusammenfassung

Nichtlinearität in Zeitreihen findet sich nicht nur in Kointegrationsbeziehungen, sondern ist auch eng verknüpft mit Strukturbrüchen und Persistenzbrüchen. Da beide Arten von Brüchen zu einem Regimewechsel führen können, eignen sich diese Modelle besonders, um Zeitabhängigkeiten zu modellieren. Lineare Modelle liefern häufig nur eine unzureichende Darstellung des vorliegenden DGPs, da sie Schocks wie Finanzkrisen, Trends oder aber stilisierte Fakten wie Volatilitätsclustering und Langzeitdependenzen nicht widerspiegeln können. Aus diesem Grund ist das Testen auf diese nichtlinearen Eigenschaften essentiell in jeder statistischen Analyse, insbesondere im Hinblick auf eine effektive Modellspezifikation.

Nichtlineare Kointegration. In Kapitel Eins wird ein nichtlinearer Kointegrationstest vorgestellt, der ursprünglich auf das Papier von Kapetanios et al. (2006) zurückgeht, in welchem erstmals nichtlineare Fehlerkorrektur explizit unter der Alternative untersucht wird. Es wird unterstellt, dass die Regimewechsel durch eine logistische Übergangsfunktion zweiter Ordnung (D-LSTR) ausgelöst werden. Getestet wird „Keine Kointegration“gegen „Global stationäre D-LSTR Kointegration„unter der Alternative des Tests. t- und F- Tests werden direkt aus der nichtlinearen Fehlerkorrekturgleichung abgeleitet und für kleine Stichprobengrößen untersucht. Die Ergebnisse des nichtlinearen Tests werden mit einem gängigen linearen Kointegrationstest verglichen, hier dem Test von Johansen (1991). Es ergibt, dass der vorgestellte nichtlineare Test Power gegen D-LSTR Kointegration, aber auch diskretes 3-Regime TAR-Verhalten hat und sich die D-LSTR Funktion sich besonders als Generalisierung der Übergangsfunktionen eignet.

Strukturbrüche. Das zweite Papier untersucht die am häufigsten angewendeten Volatilitätsbruchtests, nämlich den CUSUM-Test von Deng and Perron (2008) sowie gewöhnliche LM- und Wald-Tests. Im Rahmen einer Simulationsstudie unterliegen die DGPs entweder einem oder zwei Brüchen oder erfahren einen konstanten Anstieg in der Volatilität. Neben den üblichen Size- und Powervergleichen werden die Tests auch empirisch validiert und sowohl auf Wechselkurse als auch Aktienkursdaten angewendet. Eines der Hauptergebnisse ist, dass große Ausreißer die Langfristvarianz derartig beeinträchtigen, dass nicht-monotones Powerverhalten resultiert.

Persistenzbrüche. In Kapitel Drei wird der expliziten Frage nachgegangen, ob zusätzliche Brüche in der bedingten Volatilität dazu führen, dass der R Test von Leybourne et al. (2007) fälschlicherweise ablehnt. Da die simultan auftretenden Brüche in der bedingten Varianz nicht unter der Nullhypothese spezifiziert werden, wird erwartet, dass die Grenzverteilung nichtpivotal aufgrund von nicht identifizierten Parametern ist. Um einer nicht-pivotalen Grenzverteilung vorzubeugen, wird ein Wild Bootstrap vorgeschlagen und auf den Test von Leybourne et al. (2007) angewendet. Im Ergebnis lässt sich festhalten, dass der Test von Leybourne et al. (2007) wie erwartet Probleme mit der Differenzierung zwischen einem echten Persistenzbruch und einem Strukturbruch in der Varianz hat. Allerdings lässt sich durch die vorgeschlagene Testversion dieser Befund nicht beheben. Die empirische Studie mit Inflations- und Aktienkursdaten bestätigt die Simulationsergebnisse.

Schlagwörter: Nichtlineare Kointegration, Strukturbrüche, Persistenzbrüche

Abstract

Besides obvious nonlinear relations like a nonlinear error correction model are nonlinearities in time series closely related to structural breaks and changes in persistence. Since both kinds of changes can induce regime-switching, they qualify well to capture the characteristic of time-variability. On the contrary, linear models are often an insufficient simplification of the real underlying DGP because they fail ro reproduce trends, shocks like finance crises and stylized facts such as long-range dependencies as well as volatility clustering. This is why testing for the presence of these nonlinear properties as the first step of any statistical analysis is very crucial, especially with regard to effective model building.

Nonlinear cointegration. In the first chapter, a nonlinear cointegration test is proposed which builds on Kapetanios et al. (2006) who where the first who addressed cointegration in a nonlinear error correction framework under the alternative. The switch between regimes is modeled to follow a second order logistic smooth transition (D-LSTR) function and a null hypothesis of no cointegration is tested against globally stationary D-LSTR cointegration. From the nonlinear error correction regression, t-type and F-type statistics are derived and finite-sample investigations are conducted. The results of the modified nonlinear cointegration test are compared to a comparable linear cointegration test, namely the test proposed by Johansen (1991). The D-LSTR function qualifies well as an overall-generalization of transition functions and it is found that the D-LSTR error correction model has power against both alternatives, D-LSTR as well as 3-regime TAR nonlinearity which is nested for large γ in the D-LSTR function.

Structural breaks. The topic of the second paper is to survey the most frequently applied volatility break tests when they are employed to a broad range of different DGPs. Within a simulation study, the break tests are applied to DGPs which can exhibit either single- or double-shifting or the process can experience a smooth increase in the magnitude of the volatility break. The surveyed tests are a CUSUM test in a version proposed by Deng and Perron (2008) and conventional Wald and LM tests. Besides size and power comparisons the break tests are empirically validated and it is found that more breaks are found in equity series than in exchange rate series. One main finding is that huge outliers in the data can impact the long-run variance of the squared return process to be no longer finite which renders non-monotonic power functions.

Changes in persistence. Chapter three addresses the specific question whether either structural breaks or nonstationarity in the conditional volatility affect the testing decision of the R test proposed by Leybourne et al. (2007). The additional structural breaks in the conditional volatility process are not specified under the null hypothesis and may therefore lead to a non-pivotal limiting distribution. Hence, heteroskedasticity of an unknown form is encountered and in order to potentially robustify the testing procedure, a wild-bootstrapped version of Leybourne et al. (2007)'s R test is suggested. Within a simulation study, size and power of the originally proposed test and the wild-bootstrap analogue are compared for various constellations of simultaneous breaks in the AR parameter as well as the GARCH parameter. It is found that the Leybourne et al. (2007) test seems heavily impacted by additional structural breaks in the conditional volatility, especially in very finite sample sizes. In an empirical application the two testing procedures are applied and evaluated.

Key words: nonlinear cointegration, structural breaks, changes in persistence

Contents

Chapter 1

Introduction

Introduction

In the recent past, the subject of nonlinearity in financial econometrics has gained tremendous importance. Not only trends, evolving over the years, but also manmade shocks such as financial crises can lead to regime switches in financial time series and render the data generating processes (DGP) time-varying. Basic linear models tend to be insufficient in order to reproduce the full time series behaviour as they are incapable to incorporate regime switches, volatility clustering or long-range dependencies. Also structural breaks, as which onetime shocks are usually identified, cannot be accommodated within this model class. It is evident, that structures have become increasingly dynamic and interdependent as a result of the gradual integration between economies and financial markets. Thus, in regard of effective model building all influences that the time series data is exposed to need to be captured by the model, which is why testing for theses incidences is inevitable.

Linear models are often appealing due to their easy applicability, good communicability and interpretability and yield in many cases a comparatively fair model choice. Yet, in many other situations a linear model seems to be a rather disadvantageous simplification of the actual structure of the underlying DGP, especially in regard of effective model building. When setting up a model specification, econometricians always encounter a chance to specify the incorrect model, which is quantifiable and well-known as model risk. The importance of a valid risk management is omnipresent and it is even institutionally manifested by protection mechanisms like Basel II & III as well as Solvency II.

Protection mechanisms such as insurances or automated sell orders are considered to be a trigger for volatility clustering, which is demonstrated by the sensitivity of financial time series to external intervention. In the aftermath of the subprime crises in 2008, fund managers were obliged to implement protection mechanisms such as downside portfolio protection insurances and automated sell orders in risk management of big insurance groups for shareholders. This directive by the regulatory institutions, however, causes also reactions on financial markets: If equity markets experience a substantial drop during one trading day, this triggers automated sell orders, which further decrease the price of equities in the corresponding portfolios. Since larger portfolios are generally correlated, this could then trigger significant downward volatility. Recurrently, this exemplifies how closely incidents on financial markets and policy making decision processes are linked.

The phenomenon of volatility clustering has been popularized ever since Engle (1982) and Bollerslev (1986), respectively, introduced the conditional heteroscedasticity models ARCH and GARCH. Whenever the volatility of a time series is time varying or features clusters, the volatility process does not conform to a linear pattern. It is then well specified or forecasted by this nonlinear extension in which the volatility is modeled separately. Since its introduction, the class of ARCH/GARCH models has experienced an immense expansion and reached common practise. For his contribution to the analysis of time series data Robert Engle was awarded the Nobel Price in Economics in 2003 jointly with Clive Granger.

Granger (1981) coined the term ”cointegration“ although the pioneer work of the concept refers to Davidson et al. (1978). The phenomenon of cointegration is mostly prevalent on financial and macroeconomic markets. Two time series are deemed to be integrated, if they are individually integrated of order one while a linear combination of them yields stationarity. This concept has substantially impacted empirical applications and the concurrent perception of interdependencies on financial markets. It shed light on time series interactions and its mutual impact. Doubtlessly, this approach has shown econometricians quite plainly the importance of a thorough investigation of the patterns and dependencies of the underlying DGP. The Nobel Price committee distinctly underlined the relevance of cointegration, awarding Granger’s contributions to the analysis of both linear and nonlinear time series.

Nonlinearities are also closely related to structural breaks and persistence changes. Both kinds of changes can induce regime switching and capture the characteristic of time-variability. Clements and Hendry (1996) postulate that it should routinely be tested for structural stability because the lack of stability of the coefficients causes invalid forecasts, cf. also Antoch et al. (2019).

The topic of cointegration and its associated error correction models have been one of the most researched fields of theoretical econometrics since it emerged in the middle of the 70’s. One of the latest research branches regarding cointegration is the extension to nonlinear dynamics and regime-switching error correction mechanisms. Literature distinguishes between either time varying cointegration relations, cf. Bierens and Martins (2010), or nonlinear adjustment processes, cf. Kapetanios et al. (2006). Smooth dynamic adjustment processes enable to model unsteady and unproportional corrections of the disequilibrium error via smooth transition (STR) functions. Financial applications for nonlinear error correction can be found in price differentials or exchange rates under the purchasing power parity in the presence of transaction costs, cf. Taylor et al. (2001). Since mere nonlinear cointegration relations occur rather rarely, the nonlinear error correction models are of higher empirical relevance. Whenever short-run dynamics in the adjustment process to deviations from long-run equilibrium relations are nonlinear, nonlinear error correction models offer the best model fit, cf. Kiliç (2011).

This thesis addresses empirical techniques that are needed to analyze the nonlinear dynamics of financial markets, cf. Hsieh (1995). It is examined how statistical tests for these issues can be designed in the course of effective model building.

The first part of this thesis focusses on nonlinear cointegration, more specifically how to test for nonlinear error correction dynamics. It contributes an extension of nonlinear cointegration to the discussion of linear cointegration which has been debated for nearly 40 years. The study starts with an introduction to nonlinear cointegration and motivates its use in practice. An investigation of the finite-sample properties of the smooth transition-based cointegration test is presented, where the DGP under the alternative hypothesis is a globally stationary second order LSTR model proposed by Kapetanios et al. (2006). The therein provided procedure describes an application to long-run equilibrium relations involving real exchange rates with symmetric behavior. The properties of the double LSTR (D-LSTR) transition function are therefore utilised, since the D-LSTR features unit root behaviour within the inner regime on the one hand and symmetric behaviour in the outer regimes on the other hand. A null hypothesis of no cointegration is assumed and tested against globally stationary D-LSTR cointegration. As a matter of fact, the limiting distribution results to be non-standard due to the identification problem under the null hypothesis.

The D-LSTR function has the capability of both producing three-regime threshold AR (TAR) nonlinearity when the transition parameter tends to infinity and generates exponential-type nonlinearity that closely approximates exponential STR (ESTR) nonlinearity. For this reason, the D-LSTR function qualifies well and it is found that the D-LSTR error correction model has power against both of these alternatives.

The next two chapters deal with changes in structure and persistence, respectively, and both consider a conditional volatility process. The scope of the conditional heteroskedasticity models is in each case a GARCH(1,1). This model is well-applied by financial institutions for modeling the volatility of returns of stocks, bonds and market indices. The advantage of GARCH models is the possibility to model the expected volatility structure of returns time-varying. This allows to model the volatility structure more volatile during times of financial crises when shocks occur and less volatile during calm and steady phases of economic growth.

In chapter three, a comparative study of volatility shifts and their impact on time series in a number of testing procedures is presented. The performance of different structural break tests is evaluated in the context of various DGPs. Precisely, the size and power properties of CUSUM based, LM and Wald volatility break tests are investigated. In a simulation study, the properties of the tests are derived supposing that the processes feature shifts in both the unconditional and conditional variance whereat the DGP is exposed to single- or double-shifting or exhibits a smooth increase in the volatility. One main finding is that huge outliers in the data can impact the long-run variance of the squared return process to be no longer finite which renders non-monotonic power functions. As an empirical example, the number and timing of volatility breaks is determined considering four equity and three exchange rate series. Moreover, the distribution of the p-values is derived conducting the tests in a rolling window. The findings hereof disclose fewer volatility breaks in the currency series than in the equity series.

The last chapter picks up the foregoing discussion about structural changes and expands it to persistence changes under nonstationary conditional volatility. It is studied whether a structural change in the conditional volatility affects the testing decision of the persistence change test proposed by Leybourne et al. (2007). Cavaliere and Taylor (2008) have shown that processes which display time-varying *un*conditional volatility suffer from severe over-sizing in persistence change tests since their limiting null distributions are no longer pivotal. For this reason, it is investigated whether the variance ratio test of Leybourne et al. (2007) can be robustified against time-varying conditional volatility by applying wild bootstrap-based implementations similar to Cavaliere and Taylor (2008). The idea is to improve the capability of the test to prevent from spurious rejection of the null hypothesis when persistence changes in mean and structural changes in the volatility as well as explosive behaviour can occur simultaneously. A simulation study is conducted and the suggested procedure is empirically applied to three inflation rates and one stock index. The finding is that the bootstrapped test succeeds to detect breaks in the demeaned but not in the detrended data, whereas the original Leybourne et al. (2007) test never rejects in favour of a break in persistence.

Chapter 2

Testing for Cointegration in a Double-LSTR Framework

Testing for Cointegration in a Double-LSTR Framework

Co-authored with Philipp Sibbertsen

Published: (2015) In: Beran J., Feng Y., Hebbel H. (eds) Empirical Economic and Financial Research. Advanced Studies in Theoretical and Applied Econometrics, vol 48., pp. 437-450, Springer, Cham. https://link.springer.com

Chapter 3

A Comparative Study of Volatility Break Tests

A comparative study of volatility break tests

Co-authored with Philip Bertram

3.1 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. van Dijk et al. (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 4.3 provides a short introduction of the insinuated testing procedures. In section 3.3 the simulation study is presented and the results are discussed. Section 3.4 then contains the empirical example while section 3.5 concludes.

3.2 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

$$CUSQ = \max_{0\leq\tau\leq1} \left| \sum_{t=1}^{[\tau T]} \eta_t^2 - \sum_{t=1}^{T} \eta_t^2 \right| / \sqrt{T\Theta},$$

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 Ploberger and Krämer (1992). 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

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

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:

$$\begin{array}{ll} \text{i)} & \sup\text{-}\mathcal{J} = \sup \mathcal{J}(\tau) \\ \text{ii)} & \text{mean -}\mathcal{J} = \int \mathcal{J}(\tau) \\ \text{iii)} & \exp\text{-}\mathcal{J} = \log\left(\int \exp\left[\frac{1}{2}\mathcal{J}(\tau)\right]\right), \end{array}$$

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 $\hat{\eta}_{t-1}^2$.

3.3 Monte Carlo Study

3.3.1 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.

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

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.

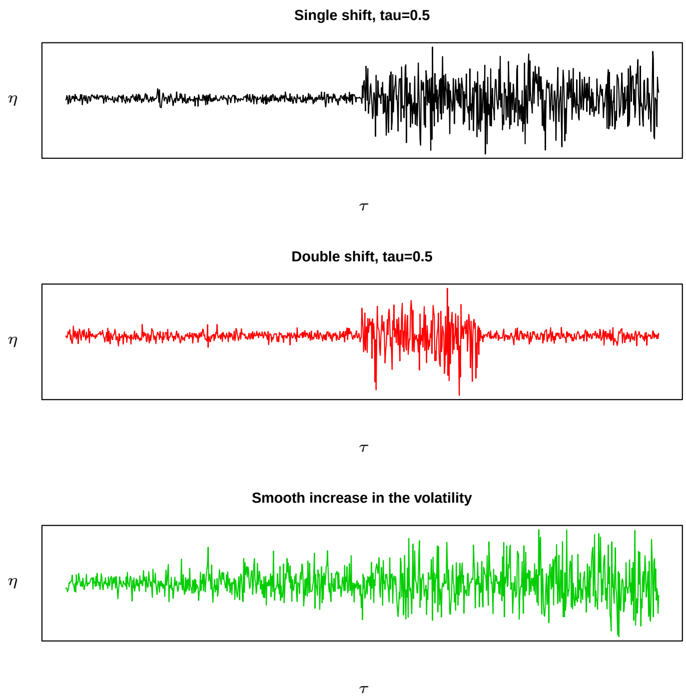

Figure 1: displays exemplary the inflation of the volatility over time for the different process types for $T = 1000$ with $\tau = 0.5$.

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

	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.

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

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

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.

3.3.2 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

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

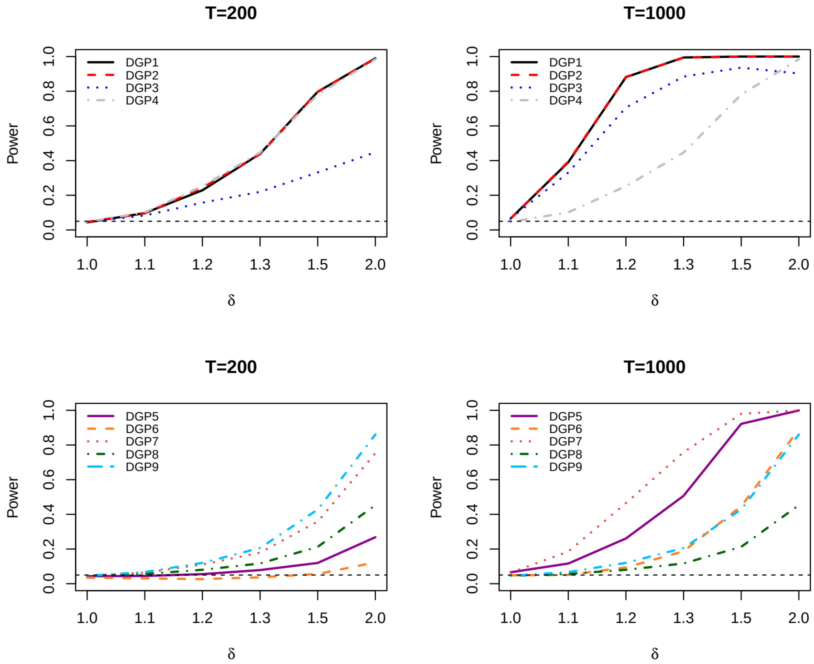

Figure 2: displays the power of the DP test for $T = 200$ (left) and $T = 1000$ (right) with $\tau = 0.5$ over all DGPs.

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 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 $\succ$ DP $\succ$ 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

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

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 with in section 3.3.3. 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 $\succ$ expLM $\succ$ 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.

3.3.3 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.* $\square$

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.

			$T = 500$		$T = 1000$		$T = 2000$	
α_1	κ	Σ	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 $\alpha_0 = 0.1$.

			$T = 500$		$T = 1000$		$T = 2000$	
α_1	κ	Σ	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 $\alpha_0 = 0.1$.

			$T = 500$		$T = 1000$		$T = 2000$	
α_1	κ	Σ	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 $\alpha_0 = 0.1$.

			$T = 500$		$T = 1000$		$T = 2000$	
α_1	κ	Σ	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.

3.3.4 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:

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

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 density for $T = 200$.

At first sight the DGPs behave qualitatively analogical over all three tests, which is why we

[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.

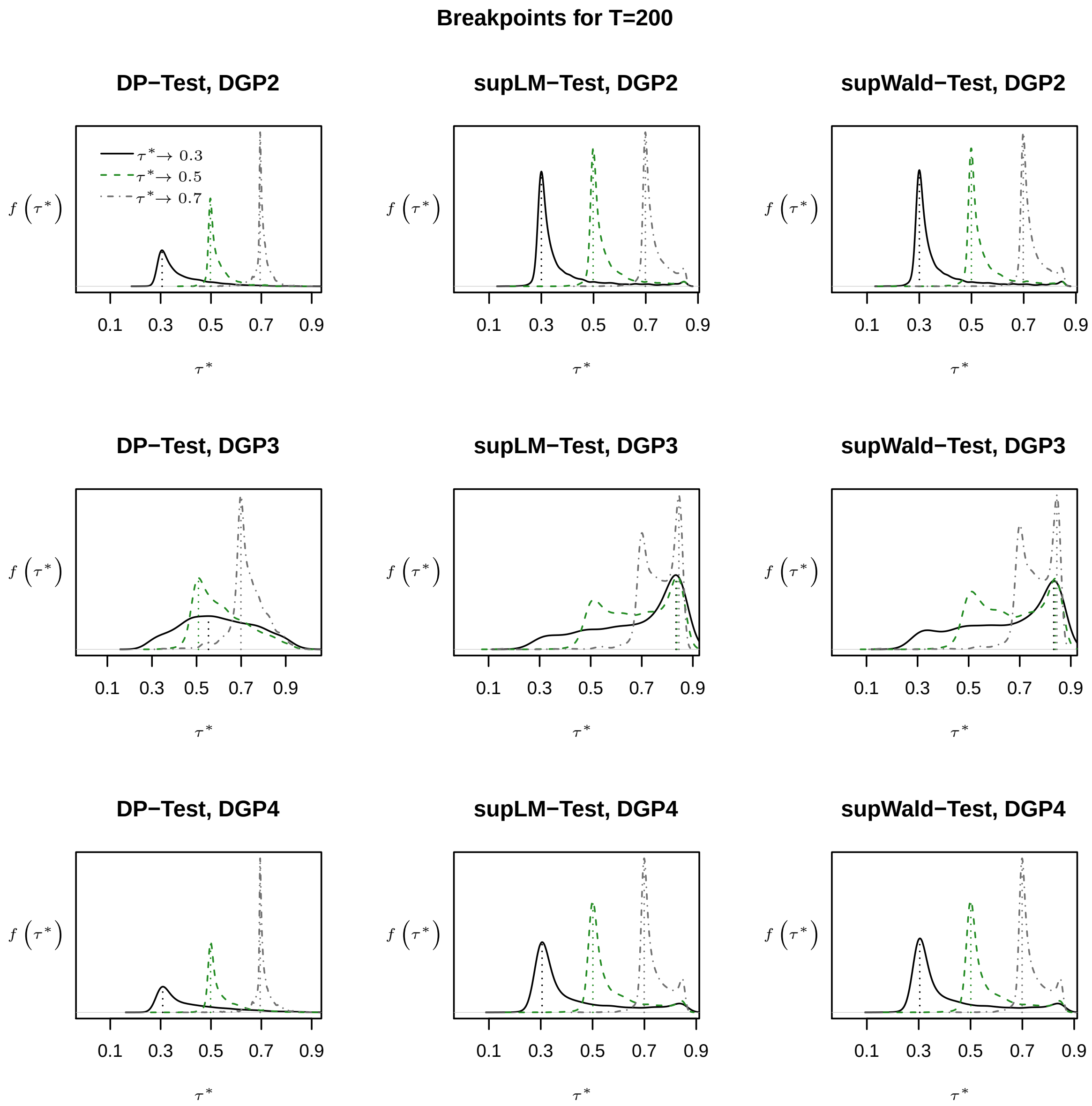

Figure 3: displays the density of the breakpoints in case of DGP 2-4 for $T = 200$, $\Sigma = 2$ and all τ.

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 case of $\tau = 0.5$ and 0.7 the density is even bimodal for the least squares tests and that the peak centers around

	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
					$T = 200$					
DGP 2	$x_{0.5}$	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
					$T = 200$					
DGP 3	$x_{0.5}$	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
					$T = 200$					
DGP 4	$x_{0.5}$	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.

$\tau = 0.85$ 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. 3.3.3.

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 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

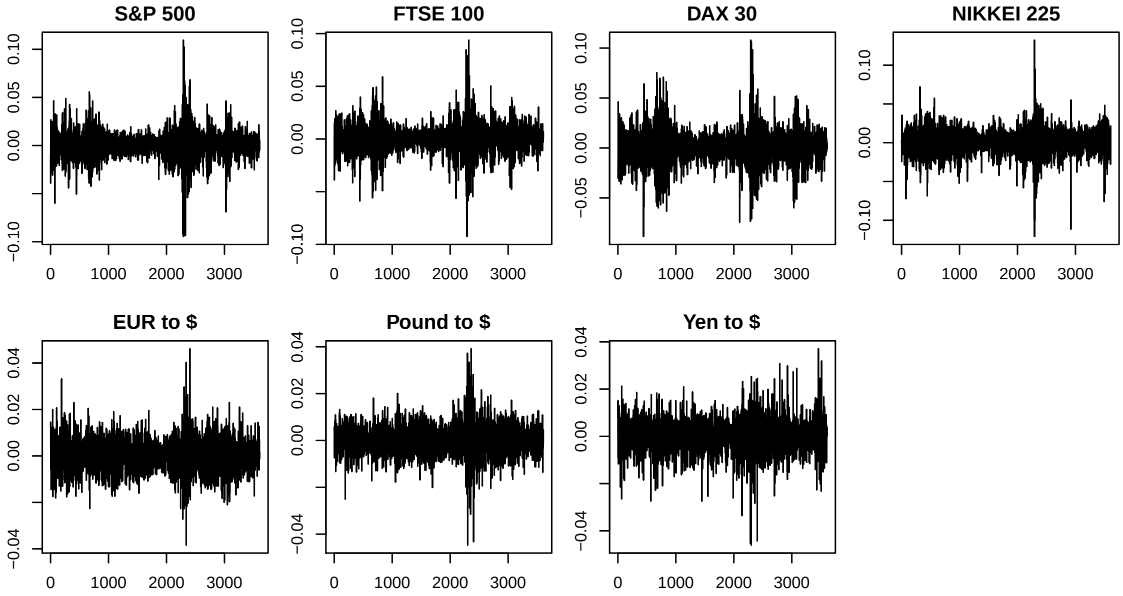

Figure 4: plots the returns of the 7 time series.

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 $\tau = 0.7$.

3.4 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, $\tau^* := \arg \max_{1 \leq \tau \leq T} \mathcal{J}_i(\tau) \mid \mathcal{J}_i(\tau^*) > Q_i^\alpha$ 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 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

	# Breaks	1	2	3	4	5	6	7
					Break Date			
					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 $\varepsilon = 0.15$.

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 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.

[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.

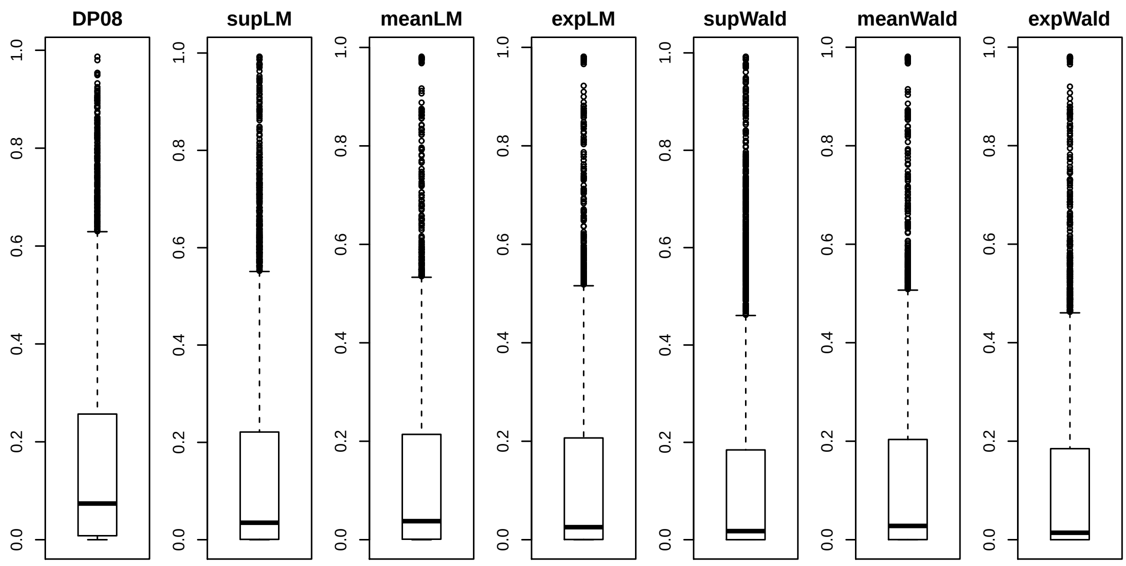

Figure 5: displays the boxplots for the p-values of the respective test on constant volatility for the S&P 500 with $\varepsilon = 0.15$.

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.

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

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

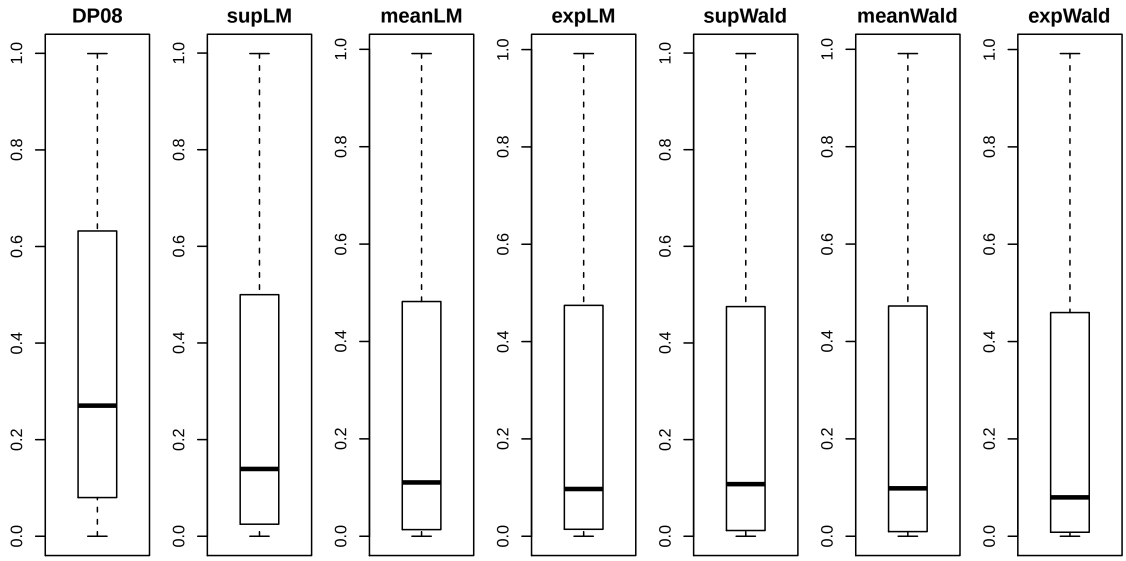

Figure 6: displays the boxplots for the p-values of the respective test on constant volatility for the \$/€ exchange rate with $\varepsilon = 0.15$.

	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.

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.

3.5 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-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. 4.4 and 3.4. 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. 3.3.4. 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.

3.6 Appendix

A Figures

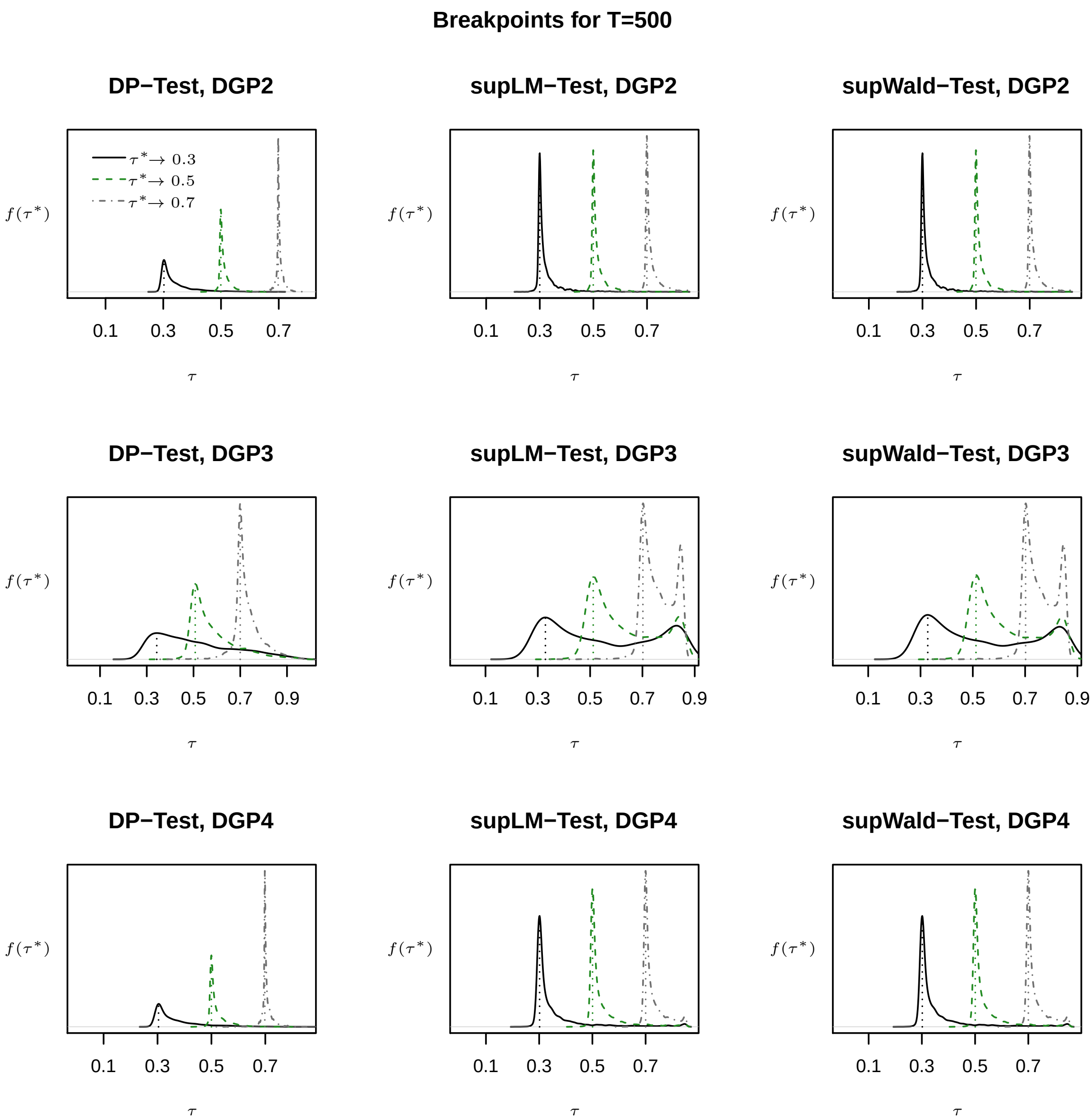

Figure 7: displays the density of the breakpoints in case of DGP 2-4 for $T = 500$, $\Sigma = 2$ and all τ.

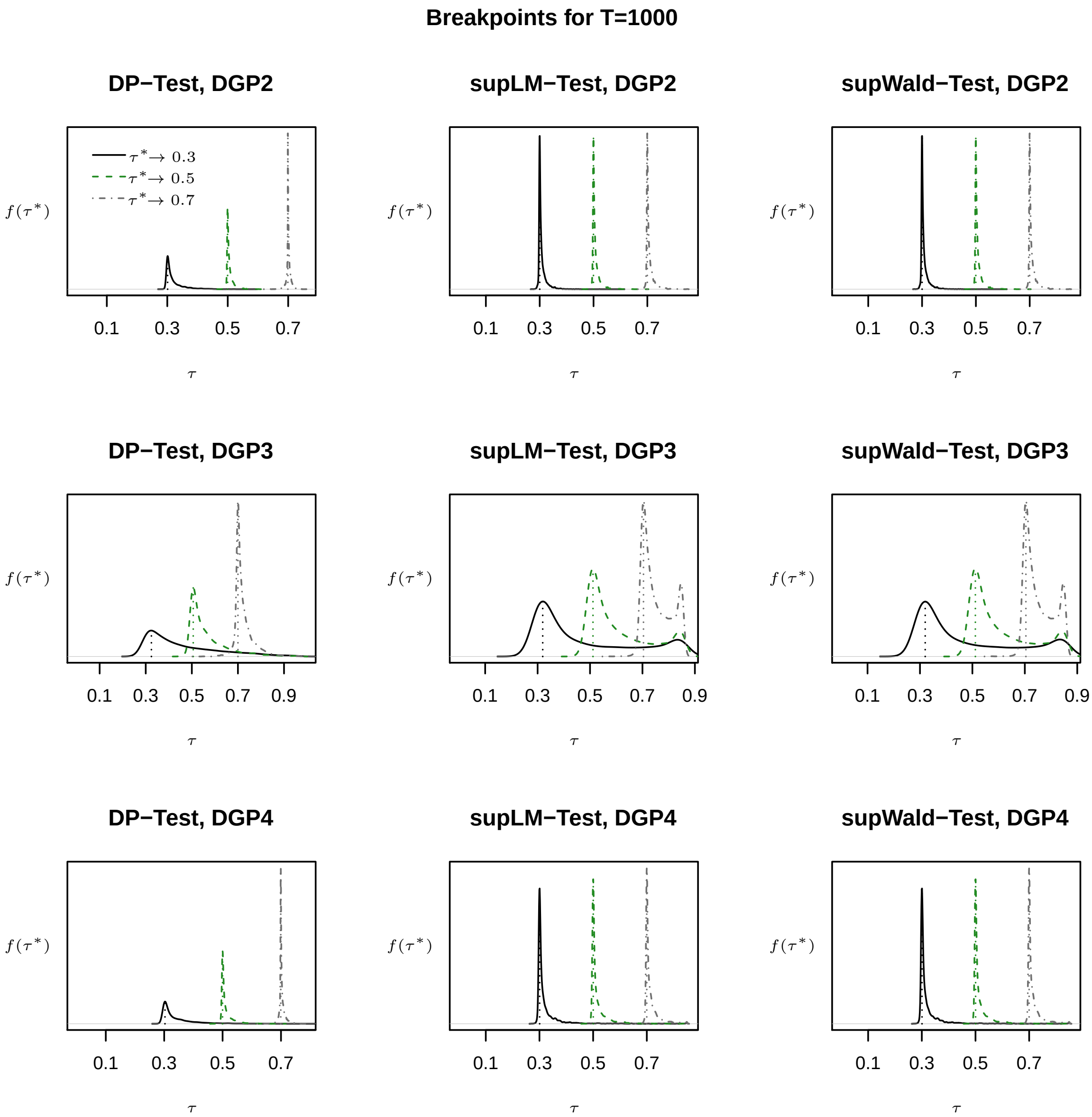

Figure 8: displays the density of the breakpoints in case of DGP 2-4 for $T = 1000$, $\Sigma = 2$ and all τ.

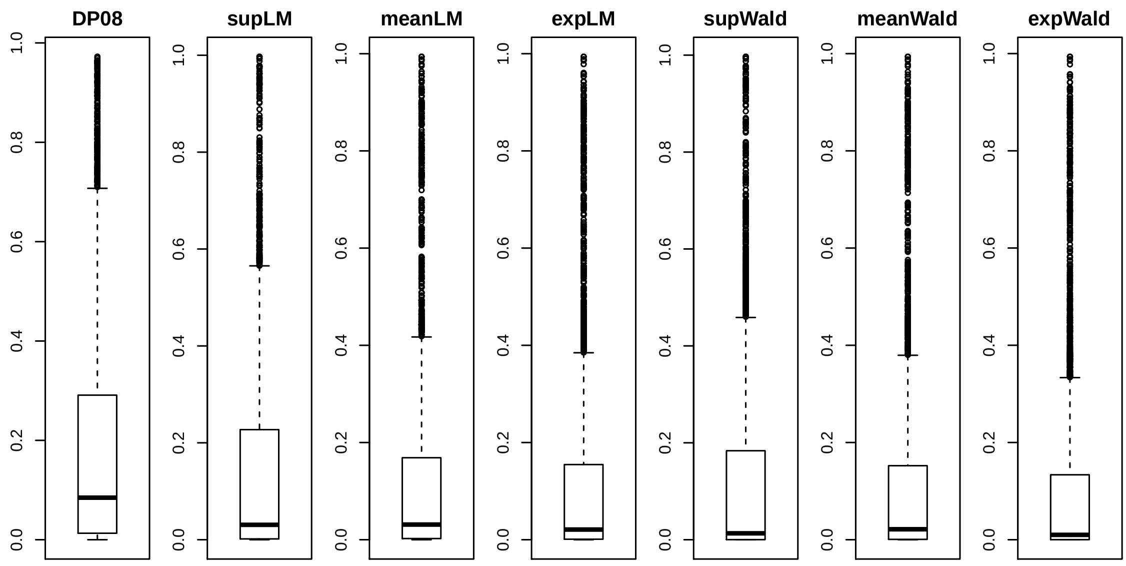

Figure 9: displays the boxplots for the p-values of the respective test on constant volatility for the FTSE 100 with $\varepsilon = 0.15$.

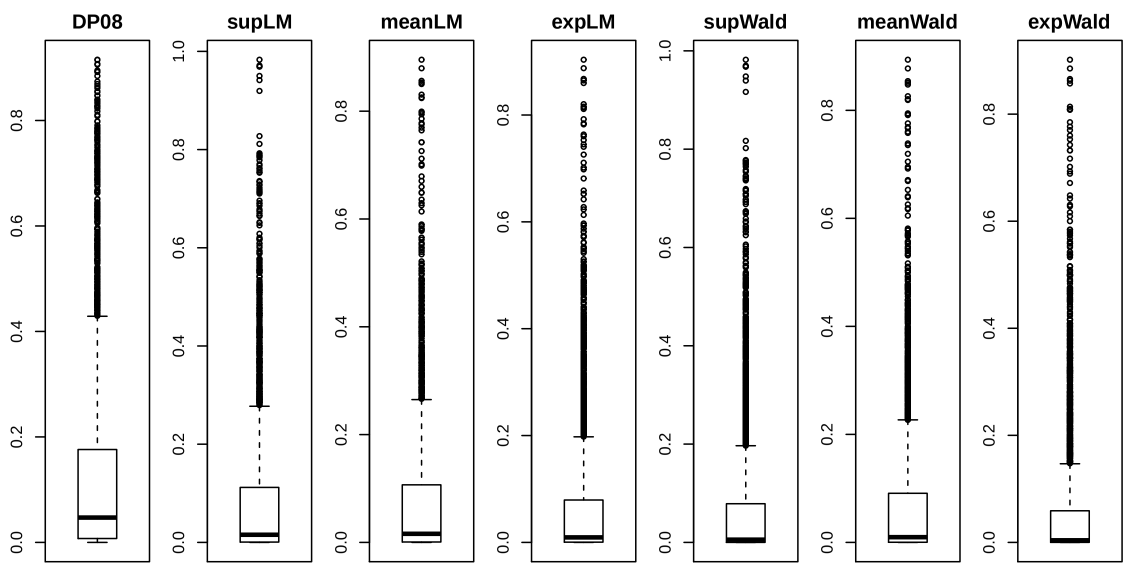

Figure 10: displays the boxplots for the p-values of the respective test on constant volatility for the DAX 30 with $\varepsilon = 0.15$.

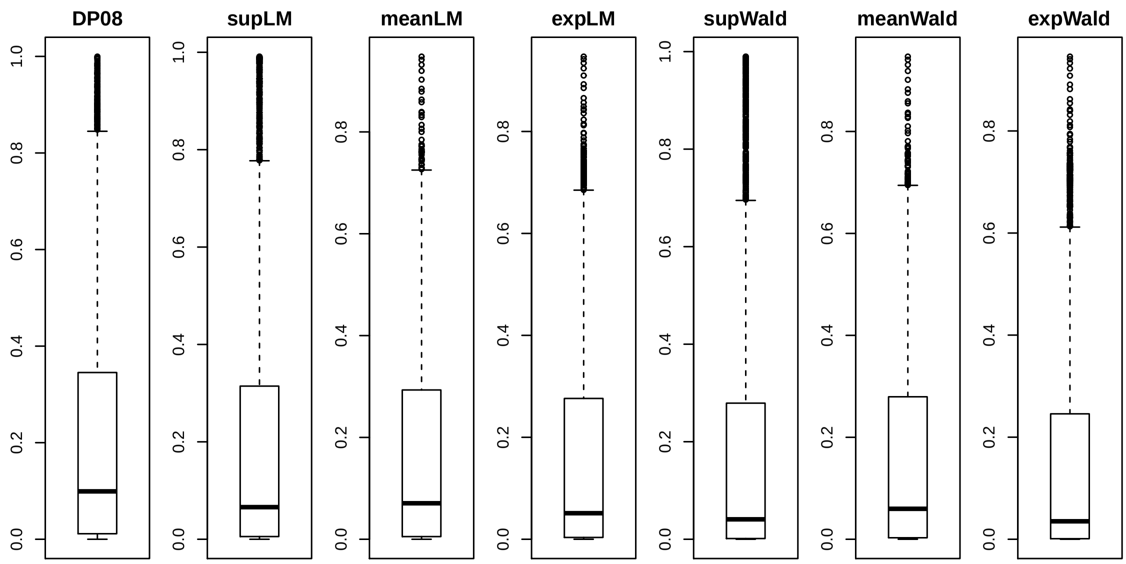

Figure 11: displays the boxplots for the p-values of the respective test on constant volatility for the Nikkei 225 with $\varepsilon = 0.15$.

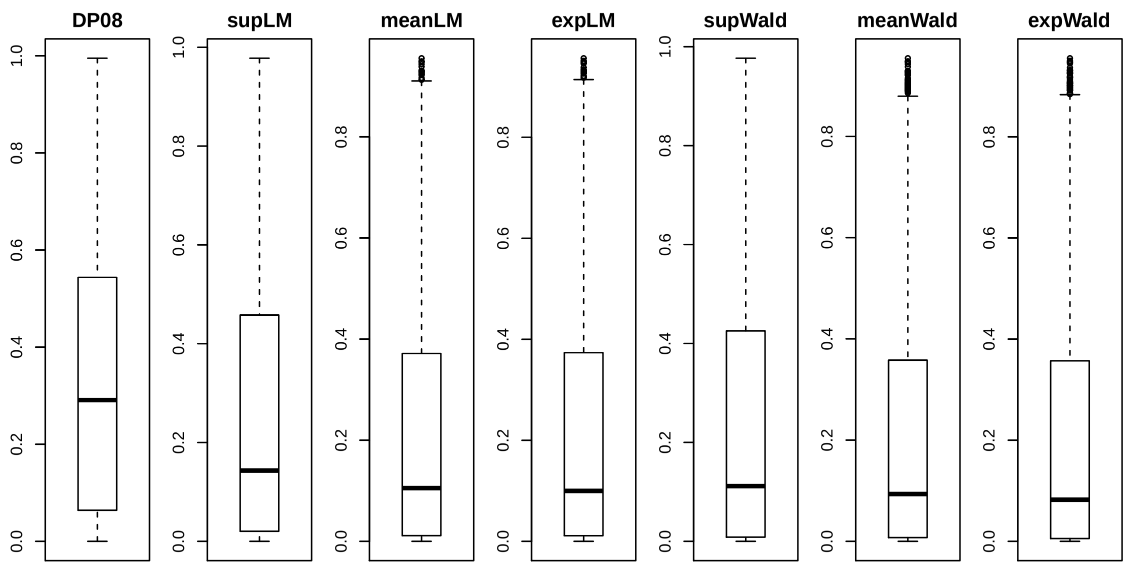

Figure 12: displays the boxplots for the p-values of the respective test on constant volatility for the Ł/\$ exchange rate with $\varepsilon = 0.15$.

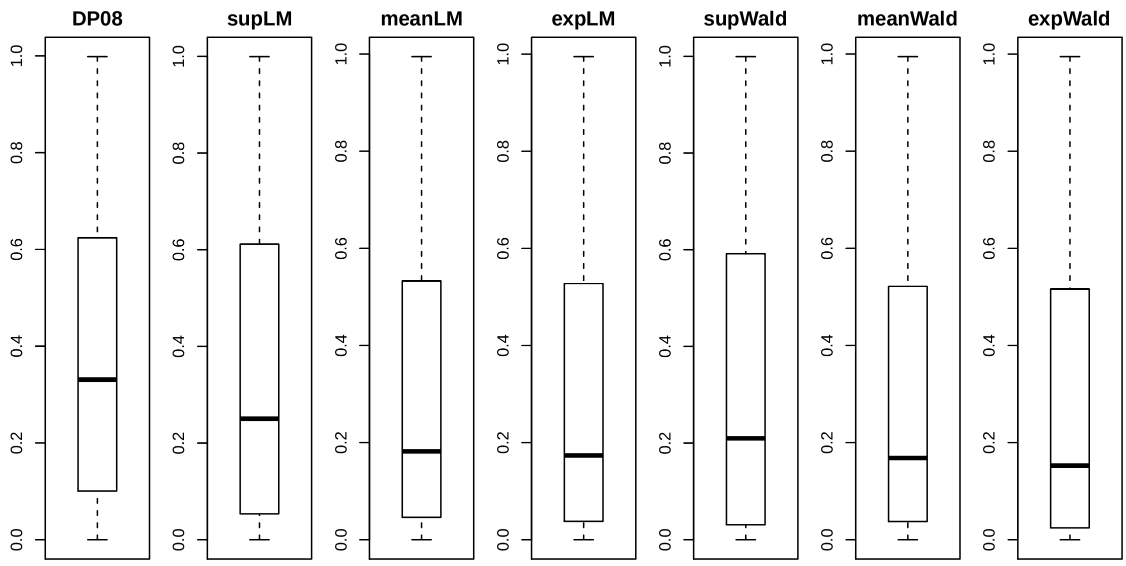

Figure 13: displays the boxplots for the p-values of the respective test on constant volatility for the ¥/$ exchange rate with $\varepsilon = 0.15$.

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 $T = 200$.

Chapter 4

Testing for Persistence Changes in the Presence of Time-varying Conditional Volatility

Testing for Persistence Changes in the Presence of Time-varying Conditional Volatility

4.1 Motivation

The detection of persistence changes has gained increasing attention, especially in regard of effective model building. Previous publications on this subject, see e.g. Busetti and Taylor (2004), Taylor (2005) or Taylor (2006), show that the ability to decompose a time series into its stationary and nonstationary components seems to be very desirable. The knowledge of the correct characteristics of the series contributes to more accurate model specifications and, hence, improves effective model building, hypothesis testing and forecasting in both applied economics and financial econometrics. Structural shifts in volatility processes have been studied for a long time. However, the current matter of interest is to determine whether the series switches from (trend) stationarity to difference stationarity. A rather recent development is the combination of volatility shifting in combination with persistence changes. Hence, the question whether either structural breaks or nonstationarity in the conditional volatility process may lead to a false rejection of the null of no persistence change constitutes the main subject of this paper.

In the matter of persistence change tests there exist two complementary approaches which differ in the behaviour of the process under the null hypothesis. The process is either of constant $I(0)$ or constant $I(1)$ behaviour and tested against changing persistence from either $I(0)$ to $I(1)$ or $I(1)$ to $I(0)$ or both, depending on the construction of the procedure. The most popular approach was suggested by Kim (2000) and Kim et al. (2002) which is designed for a null hypothesis of constant $I(0)$ behaviour. They suggest a sub-sample based ratio test of the residuals for which the idea originates from conventional stationarity tests as proposed by Nyblom and Mäkeläinen (1983) and Kwiatkowski et al. (1992), among others. The second class presumes a null hypothesis of constant $I(1)$ behaviour. Former contributions concerning the latter are based on conventional unit root testing procedures as proposed by DeJong et al. (1992a,b), for instance. The main drawback of this class concerns the fact that these tests tend to spuriously reject in the presence of the *false* null hypothesis - the behaviour which was *not* presumed under the null, namely $I(0)$. As a consequence, they suffer from low power and size properties as pointed out by Haldrup and Jansson (2006). Better power results are for instance achieved by Leybourne et al. (2003), where the procedure is derived from the Dickey Fuller unit root statistic. The most prevailing approach though is proposed by Leybourne et al. (2007), who seize the above mentioned shortcoming and introduce a test which avoids the spurious rejection in favour of a persistence change when the series features in fact $I(0)$ behaviour, even though it is conceived under $I(1)$ behavior.

Relating to volatility processes both conventional unit root tests as well as stationarity tests suffer from potentially large size distortions in the presence of nonstationary unconditional volatility as reported by Busetti and Taylor (2003) and Cavaliere (2004, 2005). Generally, persistence change tests assume unconditional volatility for the time series under consideration. In order to overcome potential size distortions Cavaliere and Taylor (2008) suggest wild bootstrap implementations for the ratio-based persistence change test proposed by Kim (2000). They counter the difficulty of detecting changes in persistence successfully by allowing nonstationarity in the unconditional volatility. Further results regarding nonstationary volatility in conjunction with changing persistence are presented in the work by Cavaliere and Taylor (2006). Contributions to dealing with structural changes in conditional volatility models refer to Hansen (2000), who employs and brings the fixed regressor bootstrap into prominence. Nevertheless, literature lacks studies on changing persistence under (non-)stationary conditional volatility.

Therefore, this paper provides an investigation on a persistence change test when the conditional volatility model of the underlying time series is subjected to structural breaks or nonstationarity. Persistence changes occur in the parameters of the AR(1), while structural breaks and nonstationarity in the volatility are induced by the scope of an GARCH(1,1), where the GARCH parameters break and can even exhibit explosive behaviour. The matter of interest is whether the properties or the behaviour of the conditional volatility impact the decision of the persistence change test. A simulation study is conducted, in which the testing procedure proposed by Leybourne et al. (2007) is combined with the fixed regressor wild bootstrap suggested in Hansen (2000) in order to account for potential size distortions due to (nonstationary) conditional volatility.

The rest of the paper is organised as follows: In section 4.2, the model setup and the conditional volatility model is outlined. Section 4.3 gives a summary of the relevant testing procedures and the testing problem of interest is motivated. Section 4.4 provides a simulation study including a discussion of results. An empirical application to inflation rates is presented in section 4.5 before section 4.6 concludes the paper.

4.2 The Model

4.2.1 Persistence Changes in Mean

The model for the following investigation has the scope of an AR(1) that contains a shift in persistence in mean and features a conditional variance, introduced subsequently in section 4.2.2. The AR(1) equals

$$y_t = \phi y_{t-1} + \varepsilon_t,$$

while the deterministic part y_{t-1} can either be a constant, $z_{t-1} = 1$, or a constant plus linear trend, $\mathbf{z}_{t-1} = (1,t)'$ with z_t denoting the data. The autoregressive parameter ϕ is dependent on τ which determines the change in persistence in mean. τ defines the change point proportion, with $\tau \in \Lambda = [\tau_L, \tau_U]$ which is an interval on $(0,1)$. Λ defines the grid for the search set later on, including the lower, τ_L, and upper, τ_U, bound. Thus, a change occurs at $\lfloor \tau T \rfloor$, whereby $\lfloor \cdot \rfloor$ denote the integer parts of its arguments. The AR-parameter ϕ switches between stationarity and nonstationarity at $\lfloor \tau T \rfloor$, meaning ϕ is $\in [0,1]$. This leads to a two-regime AR(1), c.f. section 4.2.3. Persistence changes are assumed to happen solely in the AR-parameter which is why exclusively persistence changes in mean are regarded.

4.2.2 Conditional Volatility

The family of GARCH models constitutes a class of conditional variance models, in which the variance of the process y_t is modelled dependent on both the past shocks ε_t and the past values of the variance itself, defined by σ_t. The GARCH(p,q) model extends the linear ARCH(p) model by q lags of the conditional variance, introduced in Bollerslev (1986). The scope of this work is a GARCH$(1,1)$ of the following form

$$\varepsilon_t = e_t \sigma_t \qquad \text{with}$$
$$Var\left(y_t | y_{t-1}\right) = \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta \sigma_{t-1}^2 \ ,$$

whereby $\alpha_0 > 0$, $\alpha_1 \geq 0$ and $\beta \geq 0$. It is assumed that $e_t \overset{iid}{\sim} \mathcal{N}(0,1)$ with e_t being independent of the process $\varepsilon_{t-j}, j \geq 1$. Provided that $\alpha_1 + \beta < 1$, the model is second order stationary. The unconditional variance equals $Var(\varepsilon_t) = \sigma^2 = \alpha_0/(1-\alpha_1-\beta)$, cf. Bollerslev (1986) and exists if $1-\alpha_1-\beta_1 \neq 0$. As soon as the sum of the parameters $\alpha_1+\beta \equiv \gamma$ approaches one, the effects of the past shocks on the current variance increase. Note, that only structural breaks in the conditional error term will be considered in this paper, while structural breaks in the mean of the process are disregarded. Nonstationarity in the conditional variance is induced by varying the GARCH parameters such that $\alpha_1 + \beta \geq 1$ at τT. The unconditional variance of the AR(1) remains unaffected by the model setup.

One of the simplest models in this context is the integrated GARCH (IGARCH) model suggested and investigated by Engle and Bollerslev (1986) and Lamoureux and Lastrapes (1990), respectively. That means that the lag polynomial γ of the conditional variance features a unit root and entails that shocks which impact the variance do not decay over time. Moreover, the unconditional variance does not exist (Lamoureux and Lastrapes, 1990). In this testing setup the unconditional volatility is indirectly impacted because it is investigated how inter alia explosive behaviour of the conditional volatility ($\alpha_1 + \beta \geq 1$) affects the testing decision of the persistence change test.

4.2.3 Persistence Changes under Conditional Volatility

The model can be decomposed into two regimes such as $y_t = y_{t,1} + y_{t,2}$. Both of the latter components are combined to a model with persistence changes under conditional volatility in form of a GARCH(1,1). For $i \in \{1, 2\}$ it is defined as

$$\begin{aligned} y_{t,i} &= \phi_i y_{t-1} + e_t \sigma_{t,i} \\ &\Rightarrow \begin{cases} y_{t,1} = \phi_1 y_{t-1} + e_t \left(\alpha_{0,1} + \alpha_{1,1}\varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2\right)^{0.5} & \text{for } i = 1 \\ y_{t,2} = \phi_2 y_{t-1} + e_t \left(\alpha_{0,2} + \alpha_{1,2}\varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2\right)^{0.5} & \text{for } i = 2\,. \end{cases} \end{aligned} \tag{6}$$

$y_{t,1}$ switches to $y_{t,2}$ according to the break point parameter τ. The shifting parameters, ϕ_i and $\{\alpha_{0,i}, \alpha_{1,i}, \beta_i\}$, can change simultaneously at $\lfloor \tau T \rfloor$, leading to the two-regime AR(1) process given in (6). It features either a change in persistence and/or exhibits a structural change in the conditional volatility which may lead to a nonstationary conditional volatility. Regarding the size, for instance, the $AR(1)$ remains stationary and either outcome in the conditional volatility is considered.

Four possible hypotheses can be considered employing this model setup. To be in accordance with the persistence change literature, H_0 and unconventionally H_1 denote the null hypotheses while H_{01} and H_{10} denote the alternatives. The first hypothesis is $y_t \sim I(1)$ throughout, labeled H_1. This implies unit root behaviour in the form of $\phi = 1$ for $t = 1, \ldots, T$. The second possible hypothesis includes a change from stationarity to nonstationarity, marked by H_{01}, meaning y_t is $I(0)$ and changes to $I(1)$ at the time $\lfloor \tau T \rfloor$. This, on the other hand, implies $\phi_t = \phi$, $|\phi| < 1$ for $t \leq \lfloor \tau T \rfloor$ and $\phi = 1$ for $t > \lfloor \tau T \rfloor$. The third hypothesis describes the opposite scenario of the second hypothesis, when y_t is $I(1)$ and changes to $I(0)$ at $\lfloor \tau T \rfloor$, termed H_{10}. Precisely, this means that $\phi_t = 1$ for $t \leq \lfloor \tau T \rfloor$ and $|\phi| < 1$ for $t > \lfloor \tau T \rfloor$. At last, y_t can exhibit stationary behaviour throughout which is denoted by H_0.

4.3 Tests for Changes in Persistence

The following section outlines the two most frequently applied methods as well as the suggested procedure for persistence changes from stationarity to nonstationarity or vice versa. The section starts with the description of Kim (2000) while subsequently the procedure proposed by Leybourne et al. (2007) is introduced. The first procedure accounts for nonstationary unconditional volatility and additional structural changes in the volatility, while the underlying process is subjected to changes in the persistence of the mean. The second procedure constitutes the initial test for the herein suggested method. Its test statistic is based on an augmented Dickey-Fuller-type test statistic and, therefore, complements the variance ratio test given in Kim (2000) in respect of the null hypothesis. In analogy to Cavaliere and Taylor (2008), a bootstrap-based resampling technique is suggested and implemented to Leybourne et al. (2007). The penultimate subsection motivates the testing problem and the last subsection features the applied bootstrap algorithm.

4.3.1 Kim's Variance Ratio Test

In order to test against changes in persistence, Kim (2000) proposes a residual-based test consisting of ratios of sub-sample implementations. The sub-sample implementations are derived from the KPSS test statistic for stationarity, which refers to Kwiatkowski et al. (1992). Under the null hypothesis the test assumes that the DGP is stochastically stationary around a deterministic trend. Under the alternative the DGP is tested against a change from either $I(0)$ to $I(1)$ or from $I(1)$ to $I(0)$. Since the original procedure is inconsistent, modifications have independently been brought forward by Kim et al. (2002) and Busetti and Taylor (2004). Kim et al. (2002) provide consistent tests and break point estimators under $I(0)$ to $I(1)$ changes, while the Busetti and Taylor (2004) derive locally best invariant tests against either a shift from $I(0)$ to $I(1)$ or vice versa. Generally, all of these sub-sample KPSS-type tests assume constant $I(0)$ behaviour throughout the null hypothesis. The difference between ordinary KPSS- and sub-sample KPSS-test statistics consists in the fact that the latter needs to be scaled by the long-run variance estimator, cf. Taylor (2005).

Within the work of Kim (2000), the break point $\tau \in (0, 1)$ is unknown and suspected to occur at $t = \lfloor \tau T \rfloor$. The test consists of a ratio based on the partial sums of residuals before and after the break point $\lfloor \tau T \rfloor$, which are given by

$$\begin{aligned} S_{0,t}(\tau) &\equiv \sum_{i=1}^{t} \hat{\varepsilon}_{i,\tau} \qquad \text{for } t = 1, \dots, \lfloor \tau T \rfloor \text{ and} \\ S_{1,t}(\tau) &\equiv \sum_{i=\lfloor \tau T \rfloor + 1}^{t} \tilde{\varepsilon}_{i,\tau} \quad \text{for } t = \lfloor \tau T \rfloor + 1, \dots, T, \end{aligned}$$

resulting from the ordinary least squares (OLS) regression of y_t on either an intercept, $z_t = 1$, or an intercept plus linear trend, $\mathbf{z}_t = (1, t)'$. It is emphasised that $\hat{\varepsilon}_{t,\tau}$ and $\tilde{\varepsilon}_{t,\tau}$ differ in their time horizon in order to continue the corrigendum given in Kim et al. (2002). $\hat{\varepsilon}_{t,\tau}$ represent the residuals from regressing y_t on the first $t = 1, \dots, \lfloor \tau T \rfloor$ observations,

$$\hat{\varepsilon}_{t,\tau} = y_t - \bar{y}(\tau) \text{ with } \bar{y}(\tau) = \lfloor \tau T \rfloor^{-1} \sum_{t=1}^{\lfloor \tau T \rfloor} y_t \,,$$

whereas $\tilde{\varepsilon}_{t,\tau}$ denote the OLS residuals from the above regression from $t = \lfloor \tau T \rfloor + 1, \dots, T$. Hence, the data is demeaned over $t = 1, \dots, \lfloor \tau T \rfloor$ and $t = \lfloor \tau T \rfloor + 1, \dots, T$, respectively, which leads to the residual-based ratio test statistic

$$\mathcal{K}(\tau) = \frac{(T - \lfloor \tau T \rfloor)^{-2} \sum_{t=\lfloor \tau T \rfloor + 1}^{T} S_{1,t}(\tau)^2}{\lfloor \tau T \rfloor^{-2} \sum_{t=1}^{\lfloor \tau T \rfloor} S_{0,t}(\tau)^2} \,. \tag{7}$$

Given that the change point τ^* is known, the null is rejected in favour of a change in persistence when the values of $\mathcal{K}(\tau^*)$ become large. For the more realistic case that the change point is unknown, Kim (2000) and Busetti and Taylor (2004) suggest the following three statistics based on the sequence of statistics $\{\mathcal{K}(\tau), \tau \in \Lambda\}$, which are given by

$$\begin{aligned}
\mathcal{K}_1 &:= \max_{s \in \{\lfloor \tau_L T \rfloor, \dots \lfloor \tau_U T \rfloor\}} \mathcal{K}_t\left(s/T\right); \\
\mathcal{K}_2 &:= T_*^{-1} \sum_{s=\lfloor \tau_L T \rfloor}^{\lfloor \tau_U T \rfloor} \mathcal{K}_t\left(s/T\right); \\
\mathcal{K}_3 &:= \ln \left\{ T_*^{-1} \sum_{s=\lfloor \tau_L T \rfloor}^{\lfloor \tau_U T \rfloor} \exp\left(\frac{1}{2} \mathcal{K}_t\left(s/T\right)\right) \right\},
\end{aligned}$$

whereat $T_* \equiv \lfloor \tau_L T \rfloor - \lfloor \tau_U T \rfloor + 1$ holds. The first statistic $\mathcal{K}_1$ is a Chow-type-test following Andrews (1993), which considers the maximum over the sequence, while the second statistic $\mathcal{K}_2$ was proposed by Hansen (1991) and describes a mean score statistic. The last test statistic $\mathcal{K}_3$ was suggested by Andrews and Ploberger (1994) and constitutes a mean-exponential test statistic. In each case, the null is rejected for large values of these statistics. Corrected critical values and the limiting distributions for the defined statistics for the case of constant unconditional volatility, $\sigma_t = \sigma \; \forall \, t$, are provided by Kim et al. (2002) and Busetti and Taylor (2004). More recent considerations employing appropriate long-run variance estimators for the numerator as well as the denominator can be found in Leybourne and Taylor (2004).

4.3.2 CUSUM of Squares-Based Test

Leybourne et al. (2007) employ standardized CUSUM of squared sub-sample OLS residuals in order to test against a change in persistence, notably H_{01} or H_{10}, while the null hypothesis is constructed to be of constant $I(1)$ behaviour. In comparison to the variance ratio in Kim (2000) the sum of the sub-sampled residuals is denominated by an estimator for the long-run variance (LRV) of the $\varepsilon_{t,\tau}$, $\hat{\omega}^2(\tau)$. The test statistic utilizes a ratio of OLS residuals from the forward and reversed series ($\hat{\varepsilon}^2_{t,\tau}$ and $\breve{\varepsilon}^2_{t,\tau}$), which are weighted by the corresponding LRV. Firstly, the statistic of the forward series from $t = 1, \ldots, \lfloor \tau T \rfloor$ is set up,

$$K^f(\tau) = \frac{\lfloor \tau T \rfloor^{-2} \sum_{t=1}^{\lfloor \tau T \rfloor} \hat{\varepsilon}_{t,\tau}^2}{\hat{\omega}_f^2(\tau)} \tag{8}$$

in order to test H_{01}. Accordingly, the variance estimator is calculated from

$$\hat{\omega}_f^2(\tau) = \hat{\gamma}_0 + 2 \sum_{s=1}^{m} w_{s,m} \hat{\gamma}_s \ , \quad w_{s,m} = 1 - sl^{-1} \quad \text{and}$$

$$\hat{\gamma}_s = \lfloor \tau T \rfloor^{-1} \sum_{t=1}^{\lfloor \tau T \rfloor} \Delta \hat{\varepsilon}_{t,\tau} \Delta \hat{\varepsilon}_{t-s,\tau} \ ,$$

whereat m indicates the lag truncation parameter and defines $l = m + 1$ the associated bandwidth. $w_{s,m}$ denotes a weighting coefficient that assesses proportionally higher weights to the more recent observations, while m is chosen to be $\lfloor 4(T/100)^{1/4} \rfloor$ according to Leybourne et al. (2007).

For the reversed case, the series is considered backwards, $x_t \equiv y_{T-t+1}$, so that the change is assumed to occur at time $(T - \lfloor \tau^* T \rfloor)$. By construction, the statistics K^f and K^r, which is given below, complement each other. By the same arguments a change from $I(0)$ to $I(1)$ resembles H_{10}, wherefore H_{10} is calculated from

$$K^r(\tau) = \frac{(T - \lfloor \tau T \rfloor)^{-2} \sum_{t=1}^{(T - \lfloor \tau T \rfloor)} \breve{\varepsilon}_{t,\tau}^2}{\hat{\omega}_r^2(\tau)} \tag{9}$$

by taking the reversed series x_t from $t = 1, \ldots, T - \lfloor \tau T \rfloor$. In the case of a constant, $z_t = 1$, the OLS residual from the regression of x_t on this constant $z_t = 1$ is

$$\breve{\varepsilon}_{t,\tau} = x_t - \bar{x}(1 - \tau) \qquad \text{with} \quad \bar{x}(1 - \tau) = (T - \lfloor \tau T \rfloor)^{-1} \sum_{t=1}^{\lfloor \tau T \rfloor} x_t \, .$$

Besides the OLS residuals, $\breve{\varepsilon}_{t,\tau}^2$, $K^r(\tau)$ contains also

$$\hat{\omega}_r^2(\tau) = \breve{\gamma}_0 + 2 \sum_{s=1}^{m} w_{s,m} \breve{\gamma}_s \quad \text{whereof} \quad \breve{\gamma}_s = \left(T - \lfloor \tau T \rfloor^{-1}\right) \sum_{t=1}^{(T - \lfloor \tau T \rfloor)} \Delta \breve{\varepsilon}_{t,\tau} \Delta \breve{\varepsilon}_{t-s,\tau}$$

holds. The cumulation of the sum of squared OLS residuals up to the break point $\lfloor \tau T \rfloor$ distinguishes the work of Leybourne et al. (2007) from former CUSUMs of squared residual statistics as proposed, for instance, by Brown et al. (1975) and McCabe and Harrison (1980) who utilize recursive residuals or full-sample residuals, respectively. The distinction matters in so far that K^f features different orders of magnitude for $\tau \leq \tau^*$ and $\tau > \tau^*$, resulting in the desired ability to identify H_{01} and τ^* consistently under K^f, τ^* being here the true and unknown break point percentage. In particular, $K^f(\tau)$ converges in probability to zero under H_{01} for all $\tau \leq \tau^*$, but is of order $O_p(1)$ under H_{10} for all τ. The analogue is valid for K^r, which converges in probability to zero under H_{10} for all $\tau > \tau^*$ but is of order $O_p(1)$ under H_{01} regardless of the value of τ, c.f. Leybourne et al. (2007).

Supposed that τ^* is known and H_1 is tested against a persistence change, namely H_{10} or H_{01}, then a positive by-product exists in the fact that the ratio of $R(\tau^*) = \frac{K^f(\tau^*)}{K^r(\tau*)}$ converges to infinity under H_{01} but collapses to zero for H_{10}. For this reason Leybourne et al. (2007) suggest a two-tailed test in order to test H_1 against a change in persistence, following Zivot and Andrews (2002). Assuming the change point to be unknown, this approach leads to an appropriate test statistic rejecting for either small or large values. It is given by

$$R = \frac{\inf_{\tau \in \Lambda} K^f(\tau)}{\inf_{\tau \in \Lambda} K^r(\tau)} =: \frac{N}{D}\,, \tag{10}$$

whereof the minimum of the CUSUM of squared sub-sample OLS residuals over the sequence of $\tau \in \Lambda$ of the forward series is related to its counterpart of the reversed series. The consistency of R results due to the properties of the sequences of statistics of K^f and K^r in conjunction with a two-tailed test. The proof can be found in Theorem 2 in Leybourne et al. (2007) while the limiting distribution of R which is given in Theorem 1. Notably, neither the limiting distribution of the statistics provided by Leybourne et al. (2007) nor Kim et al. (2002) depends on the long-run variance of the ε_t, ω^2. The results of the limiting distribution for R even hold without standardizing by the long-run variance ω^2. Supposed that Λ is centered around 0.5, the limiting marginal distribution of N and D is identical under H_1 (Leybourne et al., 2007).

Based on these statistics, Leybourne et al. (2007) establish a consistent break point estimator for τ^* under H_{01}, that is adapted from Busetti and Taylor (2004), and is given by

$$\hat{\tau}_K = \arg\sup_{\tau \in \Lambda} \Xi(\tau),$$

whereat

$$\Xi(\tau) = \frac{(T - \lfloor \tau T \rfloor)^{-2} \sum_{t=\lfloor \tau T \rfloor + 1}^{T} \tilde{\varepsilon}_{t,\tau}}{\lfloor \tau T \rfloor^{-2} \sum_{t=1}^{\lfloor \tau T \rfloor} \hat{\varepsilon}_{t,\tau}}\,. \tag{11}$$

Note that the denominator of Ξ equals the numerator of $K^f(\tau)$ in (8). In case of a break from $I(1)$ to $I(0)$, Busetti and Taylor (2004) analogously suggest a consistent estimator for τ^* under H_{10} by

$$\hat{\tau}_K^R = \arg\inf_{\tau \in \Lambda} \Xi(\tau).$$

4.3.3 Testing Problem

It is examined whether the utilization of wild bootstrap implementations can further robustify the test given in Leybourne et al. (2007) against structural changes in the conditional volatility in consideration of both stationary (H_0) and nonstationary (H_1) processes under the null hypothesis. According to the authors the test is robust against spuriously rejecting the false null hypothesis - *false* implies the process displays in fact $I(0)$-behaviour. As obtained by their empirical results it has to be stated that the test is rather weak in small sample sizes but gains power as T increases. The assumed model setup in this work is likely to lead to a non-pivotal limiting distribution and, therefore, may lead to size distortion of the test. Hence, two issues are surveyed. On the one hand, it is investigated if wild bootstrap implementations can lead to improved size properties, especially when the process is in fact $I(0)$ and, hence, displays the *false* behaviour. On the other hand, it is checked how well the test behaves in finite samples with $T = 100, 200$. In terms of distributions it is investigated whether a break in the second moment of the error process leads to a rejection of the null hypothesis while the test is build against a break in mean, respectively the first moment of the process.

Formal Testing Problem In this testing setup a break in persistence implies exclusively changes from $I(1)$ to $I(0)$, i.e. nonstationary to stationary - behaviour, and from $I(0)$ to $I(1)$, respectively. Fractional breaks in persistence and any other values of the long memory parameter remain disregarded. Formally, a change from $\phi_1 \in [0, 1]$ to $\phi_2 \in [1, 0]$ is assumed where ϕ_i with $i = 1, 2$ denotes the autoregressive parameters in model (6). The hypotheses are the following:

$$H_0 : \textit{No break in persistence} \quad \text{vs.} \quad H_1 : \textit{Break in persistence}$$

$$\Leftrightarrow H_0 : \phi_i < 1 \quad \forall \ i = 1, 2 \quad \text{vs.} \quad H_{01} : \phi_1 < 1 \ \textit{and} \ \phi_2 = 1 \quad \text{or}$$

$$\Leftrightarrow H_1 : \phi_i = \phi = 1 \quad \text{vs.} \quad H_{10} : \phi_1 = 1 \ \textit{and} \ \phi_2 < 1.$$

Changes in the conditional variance are not considered as part of the testing problem but are added as additional nuisance. It has to be checked whether the additional nuisance alters the outcome of the testing procedure.

4.3.4 The Wild Bootstrap Algorithm

Bootstrapping is especially known for alleviating size distortions in small samples sizes. If the null hypothesis allows conditional heteroskedasticity of unknown form, the normal bootstrap fails to imitate the behavior of the original DGP and cannot project it correctly into the bootstrap DGP. This is why wild bootstrapping qualifies, developped in Liu (1988). The following suggestions refer to Wu (1986) and Beran (1986).

Under the null hypothesis of no change in persistence, nonstationary conditional volatility is very likely to modify the limiting distribution of the test statistics in so far that they are no longer pivotal. For the case of nonstationary *un*conditional volatility, Cavaliere and Taylor (2008) proof that the residual-based test statistics introduced in 4.3.1 have no longer a pivotal limiting null distribution due to violations of the normality assumption. This causes considerably over-sized test results. Thus, it might hinder practitioners to correctly distinguish between plain persistence changes occurring in the data and structural changes in the volatility that possibly leave the volatility nonstationary. In order to overcome potential inference problems arising from a non-pivotal limiting null distribution the wild bootstrap can help. Generally, wild bootstrapping is conducted when the time series displays heteroskedastic variances.

The advantage of the wild bootstrap is that it can project the present pattern of the nonstationary volatility into the bootstrap sample which is why it is preferred over other re-sampling schemes. The pairs bootstrap e.g. can also accommodate for heteroscedasticity while the residual bootstrap is invalidated by conditional heteroscedasticity since it works only for *iid* errors. A shortcoming of the pairs bootstrap is that it does not condition on the original X matrix unlike the residual and wild bootstrap. Another option would be the block bootstrap. Besides heteroscedasticity it accounts for serial correlation what leads to a loss of efficiency when serial correlation is not present. The block bootstrap is also slightly sensitive about the right block size which is to be chosen beforehand.

As a solution to all the mentioned shortcomings the wild bootstrap was introduced. It is a non-parametric resampling technique that makes it unnecessary to specify a parametric model for the volatility process and to run a pre-test for nonstationary volatility. Moreover, the approach is robust which is why wild bootstrap-based implementations are adapted referring to Hansen (2000). Precisely, a fixed-design wild bootstrap is employed, where the regressors are treated as fixed.

In the following, the wild bootstrap algorithm is presented step-by-step in that manner it is employed in the simulation in Section 4.4. The algorithm given below is already adapted to the CUSUM of squares-based test by Leybourne et al. (2007).

The first step of the bootstrap algorithm is to generate the bootstrap DGP

$$\begin{aligned} y_t^b &= u_t^b \\ \text{with} \quad u_t^b &= f(\hat{u}_t)\eta_t \,, \qquad t = 1, \ldots, T \end{aligned} \tag{12}$$

whereby $\hat{u}_t$ defines the full sample residuals from regressing y_t on $\mathbf{x}_t$ for $t = 1, \ldots, T$. $f(\cdot)$ denotes a transformation of the residuals. Usually, this is a heteroscedasticity consistent transformation. $f(\cdot)$ is often chosen to be of the form HC_3 introduced by White et al. (1980). It has repeatedly been reported, c.f. Davidson and Flachaire (2008) e.g., that HC_3 outperforms HC_1 and in most cases HC_2 which is why it is so popular, $f(\hat{u}_i) = \hat{u}_i/(1-h_i)$, whereby the h_{ii} are the i^{th} diagonal elements of the hat matrix $H = X\left(X'X\right)^{-1}X'$. Here, the residuals remain unmodified, since the significance gain is neglectable in this scenario.

The u_t^b are multiplied by $\{\eta_t\}_{t=1}^{T}$, an independent auxiliary distribution, for instance $\mathcal{N}(0,1)$ or any other two-point distribution given in (4.3.5). The essential bootstrap residuals, $\hat{\varepsilon}_t^b$ and $\check{\varepsilon}_t^b$, are then obtained from regressing y_t^b on $\mathbf{z}_t$ via OLS. Analogously to the test statistics K^f and K^r, the sums of bootstrap residuals differ in their time horizons and are generated from utilising the last $T - \lfloor \tau T \rfloor$ and the first $\lfloor \tau T \rfloor$ observations for $\hat{\varepsilon}_t^b$ and $\check{\varepsilon}_t^b$, respectively. $\hat{\varepsilon}_t^b$ and $\check{\varepsilon}_t^b$ are mean independent of $\{y_t, \mathbf{z}_t\}_{t=1}^{T}$ and capture the pattern of heteroscedasticity of the original sample. This property, as originally noted by Wu (1986), enables the wild bootstrap to remain consistent even in the presence of heteroscedasticity or model misspecification. Based on the pseudo-residuals, K_b^f and K_b^r can be straightforwardly computed from

$$\begin{aligned} K_b^f(\tau) &= \frac{\lfloor \tau T \rfloor^{-2} \sum_{t=1}^{\lfloor \tau T \rfloor} \left(\hat{\varepsilon}_{t,\tau}^b\right)^2}{\hat{\omega}_f^2(\tau)} \quad \text{and} \\ K_b^r(\tau) &= \frac{(T - \lfloor \tau T \rfloor)^{-2} \sum_{t=1}^{(T-\lfloor \tau T \rfloor)} \left(\check{\varepsilon}_{t,\tau}^b\right)^2}{\hat{\omega}_r^2(\tau)}, \end{aligned} \tag{13}$$

as described in section 4.3.2, whereby the LRVs $\hat{\omega}_{f,b}^2(\tau)$ and $\hat{\omega}_{r,b}^2(\tau)$ are also build on the pseudo-residuals $\hat{\varepsilon}_t^b$ and $\check{\varepsilon}_t^b$. Note that y_t^b is reversed before generating the pseudo-residuals $\check{\varepsilon}_t^b$ in K_b^r.

If the true change point τ^* is known, the test statistic $R(\tau^*) = K^f(\tau^*)/K^r(\tau^*)$ is evaluated at $\tau = \tau^*$, and again diverges to positive infinity when H_1 is tested against H_{01}. However, in case of H_1 tested against a persistence change from $I(1)$ to $I(0)$, H_{10}, the statistic collapses to zero. For this reason, this test requires two tails for a possible rejection in the presence of either large or small values. In case that the change point is a priori unknown, the following bootstrap analogue is applied

$$R^b = \frac{\inf_{\tau\in\Lambda} K_b^f(\tau)}{\inf_{\tau\in\Lambda} K_b^r(\tau)} = \frac{N^b}{D^b},$$

using the minimum of the particular statistics over the search set Λ. The corresponding empirical bootstrap p-value for $R^b(\tau)$ can be computed by

$$p^b(\tau) = 1 - G^b\left(R^b(\tau); \widehat{ts}\right),$$

whereat $G^b(\cdot)$ denotes the cumulative distribution function of the bootstrap statistic R^b and $\widehat{ts}$, the empirical test statistic of the compared testing procedure.

For the performance evaluation of the bootstrap, no critical values are needed. Bootstrap p-values yield exactly the same test results and normally provide far more information. However, B, the number of bootstrap replications, needs to be chosen such that $\alpha(B+1)$ is an integer. Provided this prerequisite is met, the critical value can be estimated depending on α. In case of an upper tail test for example, the critical value is determined as the number $(1-\alpha)(B+1)$ in the (ascending) sorted list of bootstrapped test statistics R_j^b. Hence, the equal-tail bootstrap p-value is appropriate because the R test statistic rejects for either large or small values. Moreover, the distribution of the R statistics is not symmetric around zero why it is recommended to use the two-sided p-value. Correspondingly, B is chosen such that $(1-\alpha)(B+1)/2$ is an integer, leading to the general formula for the bootstrap p-value

$$p^b = 2\min\left(\frac{1}{B}\sum_{j=1}^{B} \mathbb{1}\left(ts_j^b \leq \widehat{ts}\right); \frac{1}{B}\sum_{j=1}^{B} \mathbb{1}\left(ts_j^b > \widehat{ts}.\right)\right) \tag{14}$$

A bootstrap test is said to be exact whenever $\alpha(B+1)$ is an integer and the observed statistic $\widehat{ts}$ is pivotal, meaning that the distribution does not depend on any unknown parameters, cf. MacKinnon (2009). Unfortunately, there is a proportional loss of of power of $1/B$ - the smaller the size of bootstrap samples B is chosen, the less powerful is the test, cf. Davidson and MacKinnon (2000) and Jockel (1986). The latter preferably apply $B = 9999$ as a sufficiently large number in order to approximate infinity. According to Hansen (1996), p^b is consistent for the empirical p-value $\hat{p}$ as $T \to \infty$ by standard arguments.

4.3.5 Auxiliary Two-Point Distributions

Not every distribution is suitable to serve as an auxiliary distribution, here defined as $\{\eta_t\}_{t=1}^T$. It requires certain properties like $E(\eta) = 0$ and $E(\eta^2) = 1$, which are essential for the validity of the (wild) bootstrap procedure. The pseudo-data η_t in (12) can be drawn from another distribution instead of the Gaussian distribution. The first and most popular alternative in order to gain an improved accuracy constitutes Mammen (1993), given by

$$\eta_A = \begin{cases} \frac{1+\sqrt{5}}{2} & \text{with probability } p = \frac{\sqrt{5}-1}{2\sqrt{5}} \\ \frac{1-\sqrt{5}}{2} & \text{with probability } 1-p \, . \end{cases}$$

The *Mammen distribution* fulfills convenient properties in favor of the random draws, η, to be generated from notably $E(\eta_A) = 0$, $E(\eta_A^2) = E(\eta_A^3) = 1$ and $E(\eta_A^4) = 2$. In this context, Pearson (1916) set up the following inequality which yields a desirable property for the choice of an auxiliary distribution,

$$E(\eta^4) \geq 1 + E(\eta^3)^2 \, .$$

The two-point distribution proposed by Mammen (1993) meets the desirable property even with equality.

The *Rademacher distribution* is anothter frequently applied distribution among the two-point distributions that also meets the previous property with equality as discussed by Davidson et al. (2007). The Rademacher distribution, given by

$$\eta_B = \begin{cases} -1 & \text{with probability } p = \frac{1}{2} \\ 1 & \text{with probability } 1-p \, , \end{cases}$$

features the properties $E(\eta_B) = 0$, $E(\eta_B^2) = 1$, $E(\eta_B^3) = 0$ and $E(\eta_B^4) = 1$ and further offers the possibility of higher-order improvements when the parent distribution is symmetric. Since both η_A and η_B fulfill $E(\eta_i) = 0$ and $E(\eta_i^2) = 1$ for $i = A, B$ and are independent of the full sample residuals ε_t, the (pseudo-) distribution provides consistency for the sampling distribution (Davidson et al., 2007).

4.4 Monte Carlo Study

4.4.1 Simulation setup

For the performance evaluation in finite samples the sample sizes $T = \{100, 200\}$ will be of interest. Larger sample sizes are not regarded since the bootstrap has particularly power in very finite samples and has its usage especially here. The break points are as mentioned before $\tau \in \{0.3, 0.5, 0.7\}$ and the autoregressive parameter constellations for the break in persistence will vary in $\phi \in \{0.0, 0.5, 0.9, 1\}$. The additional changes in the conditional variance can happen either in $\alpha_{1,i}$ or β_i and can be taken from Table 13, Table 14 and Table 15 containing the empirical rejection frequencies for the size and power respectively. There will also be explosive behaviour in the GARCH part, meaning that the conditional variance is not always second order stationary and the condition $\alpha_1 + \beta < 1$ is not always met. Recall that the unconditional variance remains unchanged by construction and is, therefore, supposed to be homoscedastic and stationary at any time. However, it may be impacted by the explosive behaviour of the GARCH parameters.

4.4.2 Numerical Results

Critical values were computed for the R test statistic given in (10) for the performance comparison between the bootstrapped Leybourne, in the following abbreviated by bL, and its original counterpart, in the following abbreviated by L. These are generated using pseudo-data based upon a random-walk and 100,000 Monte Carlo replications

$$y_t = y_{t-1} + \varepsilon_t\,, \quad t = 1, \ldots, T$$

with $y_0 = 0$ and $\varepsilon_t \sim \mathcal{N}(0, 1)$. In the following, Table 12 provides the lower and upper tail critical values for the R test for finite samples and larger samples in the application,
$T = \{100, 200, 500, 1000\}$. Reported are the critical values for demeaned data, $\mathbf{x}_t = 1$, as well as demeaned and detrended data, $\mathbf{x}_t = (1, t)'$ using $T = 1000$ as discrete approximation for ∞.

Table 12: Lower and upper tail critical values for the R test.

		R					
$\mathbf{x}_t$	T	0.01	0.025	0.050	0.950	0.975	0.99
1	100	0.14	0.19	0.26	3.98	5.19	7.14
	200	0.13	0.18	0.24	4.10	5.40	7.50
	500	0.13	0.17	0.23	4.28	5.69	7.93
	∞	0.13	0.18	0.23	4.27	5.72	8.01
$(1, t)'$	100	0.30	0.36	0.43	2.33	2.74	3.30
	200	0.28	0.34	0.41	2.46	2.94	3.61
	500	0.26	0.32	0.39	2.57	3.07	3.81
	∞	0.26	0.32	0.39	2.62	3.16	3.94

The simulation results suggest that both testing procedures, bL and L, perform nearly identically for $\tau = 0.3$ and $\tau = 0.7$. For efficiency reasons, the size and power results for $\tau = 0.7$ are omitted, but are available upon request. When Λ is chosen symmetrically around $\tau = 0.5$, Leybourne et al. (2007) have shown that N and D, cf. eq.(10), have identical limiting distributions. Not surprisingly, this first result is derived due to the construction of the test.

Comparing the results for the three auxiliary distributions for the bootstrap procedure, it can be be stated that none of the three auxiliary distribution outperforms the other and rather similar behaviors are observed: Even in the face of an increasing sample size, none auxiliary distribution can be identified as superior. The performance of the bootstrapped test is obviously strongly dependent on the GARCH parametrization.

For the different τ there is no distinct rule derivable. In some situations power gains are observed in others a drop, what leads to no noteworthy finding.

As a matter of fact, the size properties worsen with increasing ϕ. Comparing for instance the switch in ϕ from $0 \rightarrow 0.5$ for the corresponding first two of the three parameterizations over both samples sizes, very ambiguous results are obtained. For the stationary GARCH process, the size even gains power from $T = 100$ to $T = 200$ and for the GARCH process with explosive behaviour a drop is assessed for an increasing T.

On the contrary, the L test performs almost in all situations very presentable. Exceptionally, when the GARCH process features explosive behavior then, on the one hand, the size is not met and a small oversizing is observed. On the other hand, adverse behaviour is observable regarding the power properties, but only when the parameter ϕ switches from $1 \rightarrow 0$, the L test features very low power when the GARCH process exhibits explosive behavior. In general, it can be stated, that L shows severe power losses for switches from $\phi = 1 \rightarrow \phi = 0$ and a cesura is observable when the structural break in the conditional volatility is rendered explosive.

In regard of the power results, more specifically under the alternative H_{10}, bL shows likewise behaviour. When α breaks from $0.05 \rightarrow 0.3$, what leads to explosive behavior, bL shows comparatively low power. It is noteworthy, that he result for the same parametrization but $T = 200$ is extraordinary bad for both test. This indicates, that there must be some effects on the limiting distribution. It is possible, that this structural break in the volatility process leads to a break in the distribution of the second moment, or higher and might even render it nonmonotonic. However, it is definite that this break in α in combination with explosive behavior affects the distribution of the moments somehow. To sum this up, a further investigation of how exactly the limiting distributions of bL and L are affected is of great interest.

Table 13: Size on the 5% level for demeaned ($\mathbf{z}_t = 1$) and detrended ($\mathbf{z}_t = (1,t)'$) data of the bootstrapped Leybourne test (bL) and its original counterpart (L).

T=100									$\eta = Gaussian$				$\eta = Mammen$				$\eta = Rademacher$			
ϕ		α_0		α_1		β			$\mathbf{z}_t = 1$		$\mathbf{z}_t = (1,t)'$		$\mathbf{z}_t = 1$		$\mathbf{z}_t = (1,t)'$		$\mathbf{z}_t = 1$		$\mathbf{z}_t = (1,t)'$	
ϕ_1	ϕ_2	$\alpha_{0,1}$	$\alpha_{0,2}$	$\alpha_{1,1}$	$\alpha_{1,2}$	β_1	β_2	τ	bL	L	bL	L	bL	L	bL	L	bL	L	bL	L
0	0	1e-0.6	1e-0.6	0.5	0.5	0	0	0.3	0.017	0.000	0.022	0.000	0.016	0.000	0.015	0.000	0.029	0.000	0.030	0.000
								0.5	0.017	0.000	0.020	0.000	0.013	0.000	0.019	0.000	0.027	0.000	0.027	0.000
0	0	1e-0.6	1e-0.6	0.05	0.15	0.94	0.94	0.3	0.033	0.000	0.021	0.000	0.023	0.000	0.028	0.000	0.031	0.000	0.032	0.000
								0.5	0.033	0.000	0.033	0.000	0.024	0.000	0.025	0.000	0.028	0.000	0.031	0.001
0	0	1e-0.6	1e-0.6	0.1	0.3	0.65	0.65	0.3	0.018	0.000	0.020	0.000	0.016	0.000	0.018	0.000	0.028	0.000	0.031	0.000
								0.5	0.017	0.000	0.021	0.000	0.015	0.000	0.019	0.000	0.026	0.000	0.027	0.000
0	0	1e-0.6	1e-0.6	0.3	0.3	0	0.65	0.3	0.022	0.000	0.024	0.000	0.016	0.000	0.018	0.000	0.030	0.000	0.029	0.000
								0.5	0.024	0.000	0.026	0.000	0.015	0.000	0.018	0.000	0.031	0.000	0.033	0.000
0	0	1e-0.6	1e-0.6	0.3	0.3	0.65	0.65	0.3	0.020	0.000	0.021	0.000	0.016	0.000	0.016	0.000	0.028	0.000	0.030	0.000
								0.5	0.016	0.000	0.020	0.000	0.015	0.000	0.018	0.000	0.027	0.000	0.031	0.000
0	0	1e-0.6	1e-0.6	0.2	0.2	0.79	0.79	0.3	0.017	0.000	0.015	0.000	0.012	0.000	0.017	0.000	0.023	0.000	0.027	0.000
								0.5	0.017	0.000	0.019	0.000	0.015	0.000	0.017	0.000	0.026	0.000	0.027	0.000
0	0	1e-0.6	1e-0.6	0.2	0.2	0.79	0.89	0.3	0.032	0.000	0.032	0.001	0.019	0.000	0.022	0.000	0.032	0.000	0.031	0.001
								0.5	0.029	0.000	0.031	0.001	0.021	0.000	0.023	0.000	0.026	0.000	0.033	0.001
0	0.5	1e-0.6	1e-0.6	0.1	0.2	0.4	0.4	0.3	0.354	0.000	0.383	0.000	0.331	0.000	0.371	0.001	0.423	0.000	0.450	0.000
								0.5	0.353	0.000	0.376	0.000	0.328	0.000	0.367	0.001	0.426	0.000	0.451	0.001
0	0.5	1e-0.6	1e-0.6	0.05	0.15	0.94	0.94	0.3	0.262	0.001	0.283	0.020	0.220	0.001	0.240	0.020	0.245	0.001	0.267	0.023
								0.5	0.268	0.001	0.280	0.021	0.223	0.001	0.245	0.022	0.235	0.001	0.260	0.021
0	0.5	1e-0.6	1e-0.6	0.1	0.3	0.65	0.65	0.3	0.372	0.001	0.386	0.004	0.347	0.000	0.375	0.004	0.406	0.001	0.439	0.004
								0.5	0.362	0.000	0.386	0.004	0.337	0.000	0.382	0.005	0.417	0.000	0.428	0.004
0.9	0.9	1e-0.6	1e-0.6	0.05	0.05	0.94	0.94	0.3	0.551	0.001	0.512	0.020	0.500	0.001	0.528	0.019	0.538	0.002	0.560	0.020
								0.5	0.518	0.002	0.517	0.018	0.504	0.001	0.518	0.022	0.526	0.001	0.546	0.021
0.9	0.9	1e-0.6	1e-0.6	0.1	0.2	0.69	0.79	0.3	0.486	0.006	0.474	0.038	0.463	0.005	0.458	0.040	0.481	0.005	0.477	0.038
								0.5	0.486	0.006	0.463	0.038	0.459	0.006	0.453	0.041	0.473	0.004	0.485	0.043
1	1	1e-0.6	1e-0.6	0.3	0.3	0.65	0.65	0.3	0.651	0.056	0.605	0.076	0.646	0.055	0.593	0.074	0.669	0.060	0.630	0.081
								0.5	0.661	0.060	0.596	0.074	0.647	0.058	0.591	0.074	0.668	0.059	0.620	0.070
1	1	1e-0.6	1e-0.6	0.1	0.3	0.45	0.65	0.3	0.643	0.062	0.585	0.079	0.631	0.059	0.561	0.074	0.644	0.061	0.592	0.077
								0.5	0.646	0.058	0.580	0.076	0.640	0.061	0.567	0.077	0.643	0.064	0.591	0.080
1	1	1e-0.6	1e-0.6	0.2	0.2	0.79	0.89	0.3	0.589	0.082	0.498	0.111	0.556	0.080	0.474	0.110	0.555	0.081	0.494	0.110
								0.5	0.580	0.078	0.497	0.109	0.557	0.080	0.476	0.111	0.569	0.075	0.496	0.105

Results obtained with $M = 10,000$ replications and $B = 999$ bootstrap samples and additional random noise η for the bootstrap.

Table 14: Size on the 5% level for demeaned ($\mathbf{z}_t = 1$) and detrended ($\mathbf{z}_t = (1, t)'$) data of the bootstrapped Leybourne test (bL) and its original counterpart (L).

T=200									$\eta = Gaussian$				$\eta = Mammen$				$\eta = Rademacher$			
ϕ		α_0		α_1		β			$\mathbf{z}_t = 1$		$\mathbf{z}_t = (1,t)'$		$\mathbf{z}_t = 1$		$\mathbf{z}_t = (1,t)'$		$\mathbf{z}_t = 1$		$\mathbf{z}_t = (1,t)'$	
ϕ_1	ϕ_2	$\alpha_{0,1}$	$\alpha_{0,2}$	$\alpha_{1,1}$	$\alpha_{1,2}$	β_1	β_2	τ	bL	L	bL	L	bL	L	bL	L	bL	L	bL	L
0	0	1e-0.6	1e-0.6	0.5	0.5	0	0	0.3	0.018	0.000	0.021	0.000	0.015	0.000	0.018	0.000	0.026	0.000	0.029	0.000
								0.5	0.019	0.000	0.018	0.000	0.015	0.000	0.015	0.000	0.031	0.000	0.028	0.000
0	0	1e-0.6	1e-0.6	0.05	0.15	0.94	0.94	0.3	0.038	0.000	0.042	0.001	0.028	0.000	0.033	0.001	0.031	0.000	0.033	0.001
								0.5	0.034	0.000	0.037	0.001	0.028	0.000	0.027	0.001	0.032	0.000	0.030	0.001
0	0	1e-0.6	1e-0.6	0.1	0.3	0.65	0.65	0.3	0.018	0.000	0.021	0.000	0.018	0.000	0.017	0.000	0.028	0.000	0.025	0.000
								0.5	0.021	0.000	0.022	0.000	0.018	0.000	0.018	0.000	0.024	0.000	0.027	0.000
0	0	1e-0.6	1e-0.6	0.3	0.3	0	0.65	0.3	0.022	0.000	0.025	0.000	0.021	0.000	0.020	0.000	0.027	0.000	0.030	0.000
								0.5	0.022	0.000	0.025	0.000	0.018	0.000	0.018	0.000	0.027	0.000	0.029	0.000
0	0	1e-0.6	1e-0.6	0.3	0.3	0.65	0.65	0.3	0.019	0.000	0.019	0.000	0.015	0.000	0.018	0.000	0.030	0.000	0.030	0.000
								0.5	0.019	0.000	0.019	0.000	0.015	0.000	0.019	0.000	0.026	0.000	0.030	0.000
0	0	1e-0.6	1e-0.6	0.2	0.2	0.79	0.79	0.3	0.019	0.000	0.018	0.000	0.015	0.000	0.016	0.000	0.025	0.000	0.030	0.000
								0.5	0.017	0.000	0.019	0.000	0.017	0.000	0.017	0.000	0.027	0.000	0.026	0.000
0	0	1e-0.6	1e-0.6	0.2	0.2	0.79	0.89	0.3	0.035	0.000	0.036	0.000	0.025	0.000	0.031	0.001	0.027	0.000	0.031	0.001
								0.5	0.036	0.000	0.037	0.001	0.028	0.000	0.028	0.001	0.029	0.000	0.034	0.001
0	0.5	1e-0.6	1e-0.6	0.1	0.2	0.4	0.4	0.3	0.571	0.000	0.591	0.000	0.546	0.000	0.589	0.000	0.641	0.000	0.027	0.000
								0.5	0.538	0.000	0.591	0.000	0.548	0.000	0.590	0.000	0.642	0.000	0.661	0.000
0	0.5	1e-0.6	1e-0.6	0.05	0.15	0.94	0.94	0.3	0.233	0.001	0.241	0.020	0.178	0.001	0.191	0.019	0.180	0.000	0.205	0.020
								0.5	0.230	0.001	0.235	0.018	0.180	0.001	0.189	0.019	0.190	0.000	0.196	0.022
0	0.5	1e-0.6	1e-0.6	0.1	0.3	0.65	0.65	0.3	0.531	0.001	0.547	0.001	0.507	0.000	0.523	0.001	0.563	0.000	0.590	0.001
								0.5	0.527	0.000	0.546	0.001	0.494	0.000	0.523	0.001	0.558	0.000	0.577	0.001
0.9	0.9	1e-0.6	1e-0.6	0.05	0.05	0.94	0.94	0.3	0.514	0.000	0.544	0.003	0.511	0.000	0.545	0.003	0.535	0.001	0.569	0.002
								0.5	0.516	0.002	0.548	0.003	0.504	0.000	0.541	0.003	0.535	0.001	0.574	0.004
0.9	0.9	1e-0.6	1e-0.6	0.1	0.2	0.69	0.79	0.3	0.510	0.002	0.525	0.023	0.491	0.001	0.510	0.025	0.503	0.001	0.525	0.023
								0.5	0.502	0.002	0.523	0.021	0.494	0.001	0.509	0.021	0.506	0.002	0.530	0.022
1	1	1e-0.6	1e-0.6	0.3	0.3	0.65	0.65	0.3	0.769	0.060	0.739	0.082	0.770	0.065	0.735	0.080	0.770	0.060	0.740	0.083
								0.5	0.773	0.061	0.742	0.081	0.764	0.060	0.732	0.086	0.775	0.060	0.743	0.080
1	1	1e-0.6	1e-0.6	0.1	0.3	0.45	0.65	0.3	0.769	0.063	0.712	0.084	0.755	0.060	0.710	0.083	0.760	0.060	0.722	0.081
								0.5	0.757	0.063	0.715	0.082	0.758	0.060	0.707	0.083	0.764	0.062	0.728	0.077
1	1	1e-0.6	1e-0.6	0.2	0.2	0.79	0.89	0.3	0.659	0.120	0.586	0.178	0.623	0.115	0.581	0.184	0.614	0.115	0.572	0.179
								0.5	0.648	0.108	0.600	0.184	0.630	0.118	0.574	0.184	0.627	0.113	0.726	0.074

Results obtained with $M = 10,000$ replications and $B = 999$ bootstrap samples and additional random noise η for the bootstrap.

Table 15: Empirical rejection frequencies on the 5% level for demeaned ($\mathbf{z}_t = 1$) and detrended ($\mathbf{z}_t = (1,t)'$) data.

T=100									$\eta = Gaussian$				$\eta = Mammen$				$\eta = Rademacher$			
ϕ		α_0		α_1		β			$\mathbf{z}_t = 1$		$\mathbf{z}_t = (1,t)'$		$\mathbf{z}_t = 1$		$\mathbf{z}_t = (1,t)'$		$\mathbf{z}_t = 1$		$\mathbf{z}_t = (1,t)'$	
ϕ_1	ϕ_2	$\alpha_{0,1}$	$\alpha_{0,2}$	$\alpha_{1,1}$	$\alpha_{1,2}$	β_1	β_2	τ	bL	L	bL	L	bL	L	bL	L	bL	L	bL	L
0	1	1e-0.6	1e-0.6	0.05	0.05	0.94	0.94	0.3	0.995	0.505	0.981	0.539	0.996	0.500	0.982	0.546	0.997	0.487	0.988	0.540
								0.5	0.995	0.497	0.980	0.545	0.995	0.492	0.982	0.546	0.997	0.496	0.987	0.547
0	1	1e-0.6	1e-0.6	0.05	0.3	0.94	0.94	0.3	0.830	0.490	0.797	0.646	0.771	0.483	0.750	0.651	0.776	0.487	0.782	0.654
								0.5	0.834	0.488	0.803	0.648	0.772	0.490	0.755	0.655	0.781	0.494	0.775	0.645
0	1	1e-0.6	1e-0.6	0.1	0.3	0.45	0.65	0.3	0.995	0.595	0.989	0.715	0.993	0.591	0.986	0.709	0.995	0.598	0.989	0.711
								0.5	0.994	0.608	0.986	0.710	0.995	0.593	0.980	0.663	0.995	0.603	0.987	0.704
0	1	1e-0.6	1e-0.6	0.2	0.2	0.79	0.89	0.3	0.978	0.597	0.962	0.721	0.972	0.603	0.951	0.718	0.976	0.612	0.965	0.717
								0.5	0.978	0.606	0.961	0.720	0.972	0.591	0.954	0.727	0.976	0.595	0.960	0.723
1	0	1e-0.6	1e-0.6	0.05	0.05	0.94	0.94	0.3	0.998	0.332	0.993	0.461	0.998	0.319	0.993	0.457	0.999	0.326	0.996	0.459
								0.5	0.998	0.325	0.993	0.458	0.998	0.324	0.994	0.456	0.999	0.333	0.995	0.459
1	0	1e-0.6	1e-0.6	0.05	0.3	0.94	0.94	0.3	0.666	0.085	0.523	0.181	0.652	0.081	0.523	0.191	0.664	0.087	0.528	0.189
								0.5	0.660	0.083	0.528	0.185	0.649	0.080	0.510	0.183	0.664	0.081	0.531	0.183
1	0	1e-0.6	1e-0.6	0.1	0.3	0.45	0.65	0.3	0.979	0.195	0.942	0.277	0.979	0.196	0.941	0.276	0.979	0.195	0.946	0.272
								0.5	0.977	0.194	0.940	0.278	0.980	0.194	0.928	0.281	0.980	0.200	0.954	0.285
1	0	1e-0.6	1e-0.6	0.2	0.2	0.79	0.89	0.3	0.940	0.183	0.900	0.287	0.948	0.184	0.893	0.291	0.953	0.185	0.907	0.285
								0.5	0.952	0.182	0.904	0.285	0.952	0.191	0.900	0.290	0.952	0.180	0.904	0.295
T=200																				
0	1	1e-0.6	1e-0.6	0.05	0.05	0.94	0.94	0.3	1.000	0.872	1.000	0.872	1.000	0.788	1.000	0.871	1.000	0.783	1.000	0.869
								0.5	1.000	0.785	1.000	0.875	1.000	0.780	1.000	0.865	1.000	0.785	1.000	0.875
0	1	1e-0.6	1e-0.6	0.05	0.3	0.94	0.94	0.3	0.828	0.500	0.820	0.710	0.765	0.498	0.776	0.715	0.775	0.500	0.794	0.717
								0.5	0.836	0.500	0.830	0.714	0.757	0.490	0.769	0.709	0.778	0.500	0.789	0.787
0	1	1e-0.6	1e-0.6	0.1	0.3	0.45	0.65	0.3	1.000	0.870	1.000	0.956	1.000	0.863	1.000	0.957	1.000	0.863	1.000	0.952
								0.5	1.000	0.866	1.000	0.956	1.000	0.870	1.000	0.956	1.000	0.864	1.000	0.952
0	1	1e-0.6	1e-0.6	0.2	0.2	0.79	0.89	0.3	0.993	0.789	0.987	0.914	0.989	0.789	0.985	0.916	0.987	0.794	0.988	0.911
								0.5	0.991	0.790	0.987	0.913	0.988	0.794	0.984	0.911	0.988	0.787	0.986	0.913
1	0	1e-0.6	1e-0.6	0.05	0.05	0.94	0.94	0.3	1.000	0.781	1.000	0.928	1.000	0.771	1.000	0.926	1.000	0.787	1.000	0.932
								0.5	1.000	0.768	1.000	0.929	1.000	0.780	1.000	0.928	1.000	0.781	1.000	0.928
1	0	1e-0.6	1e-0.6	0.05	0.3	0.94	0.94	0.3	0.173	0.010	0.128	0.021	0.152	0.006	0.116	0.023	0.170	0.009	0.127	0.024
								0.5	0.168	0.008	0.125	0.022	0.158	0.008	0.124	0.022	0.178	0.008	0.124	0.023
1	0	1e-0.6	1e-0.6	0.1	0.3	0.45	0.65	0.3	0.995	0.543	0.984	0.673	0.995	0.556	0.987	0.674	0.996	0.534	0.987	0.674
								0.5	0.996	0.546	0.985	0.672	0.996	0.543	0.986	0.676	0.997	0.530	0.987	0.677
1	0	1e-0.6	1e-0.6	0.2	0.2	0.79	0.89	0.3	0.895	0.340	0.834	0.485	0.899	0.341	0.835	0.483	0.896	0.336	0.833	0.485
								0.5	0.901	0.347	0.834	0.494	0.890	0.343	0.835	0.495	0.901	0.341	0.835	0.482

Results obtained with $M = 10,000$ replications and $B = 999$ bootstrap samples and additional random noise η for the bootstrap.

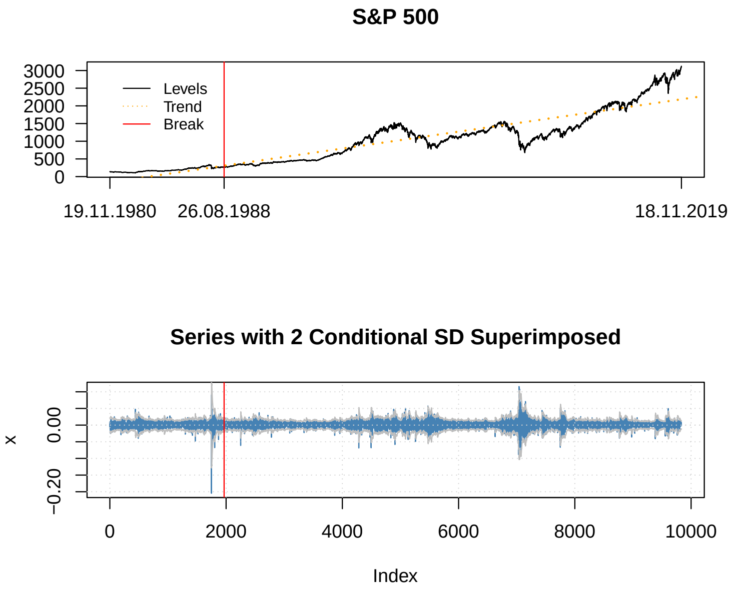

Figure 14: Plots of the demeaned time series and its GARCH fit.

4.5 Empirical Application

4.5.1 Model Fitting

As an empirical example four financial time series are considered, specifically three inflation rates and one stock data index. The stock data index, namely the S&P 500, is taken from *Yahoo Finance* and shows no seasonality and is on a daily basis. The three inflation rates are obtained from the homepage of the Federal Reserve Bank of St. Louis, the department of Economic Research, called *FRED*. Two Consumer Price Indices (CPI) are obtained: the CPI for all Urban Consumers: New Vehicles in US City Average, in the following abbreviated by *CPI Newvec*, and the CPI for all Urban Consumers: Rent of Primary Residence in US City Average, in the following abbreviated by *CPI Rent*. The third inflation rate is the real personal consumption expenditure excluding food and energy, in the following abbreviated by *Consumption*. The inflation rates are all seasonally adjusted, normed to $1982 - 1984 = 100$ and on a monthly basis. Since the series are only indirectly comparable, their lengths are deliberately chosen differently, which is summed in Table 16.

Table 16: Sums the information about the data.

Name	Basis	Start	End	# obs
CPI Newvec	monthly	02/01/1953	10/01/2019	802
CPI Rent	monthly	01/01/1981	11/01/2019	466
Consumption	monthly	02/01/2002	10/01/2019	214
S&P 500	daily	11/19/1980	11/18/2019	9835

All series are measured as the first difference of the logarithm of its level values in order to obtain stationarity. In the next step, a GARCH model is fitted. In accordance with the simulation set up in Section 4.4, a GARCH(1,1) is predefined and the best ARMA(p,q) is determined on the basis of the lowest AIC. Thus, in addition to the presumed AR(1)-GARCH(1,1) in Section

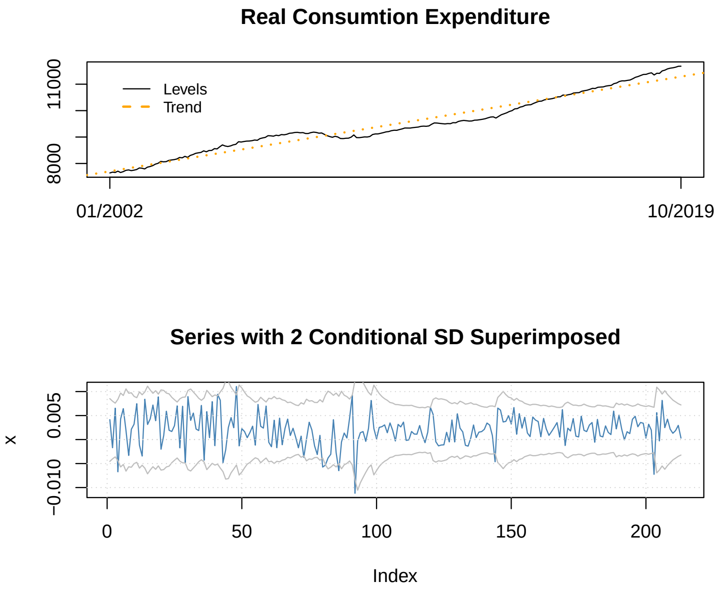

Figure 15: Plots of the demeaned time series and its GARCH fit.

4.4 also the best ARMA(p,q)-GARCH(1,1) is fitted for a comparison.

As a first step of analyzing the data, the best ARIMA(p,d,q) model is searched and validated by the lowest AIC. Whether the time series exhibits conditional heteroscedasticity or not is checked by the investigation of the ACF plot of the residuals from the best ARIMA fit. If the residuals of the fitted values seem to be a realization of a White Noise process, the ACF plot of the squared residuals is examined. If the ACF plot of the squared residuals shows conditional behaviour, meaning enduring correlated lags, a GARCH(1,1) for the conditional variance process is fitted in accordance to the approach in Section 4.4. All four series show conditional heteroscedasticity. Hence, for all four series the best ARMA(p,q)-GARCH(1,1) is fitted as well as the presumed AR(1)-GARCH(1,1).

This procedure is applied to demeaned as well as detrended data. Note that, the ARIMA(p,d,q) indicates that no further differencing is needed for the demeaned data in order to render them stationary. For the detrended series the ARIMA(p,d,q) sometimes suggests a $d = 1$, but no further differencing is applied here. After the GARCH check and the model specification, the fitted models are applied to the bootstrapped Leybourne test and its original counterpart. Table 17 summarizes the model fits and test results for demeaned data and Table 18 for detrended data, respectively. Furthermore, information about an existing trend are provided for the demeaned data. For all four series a linear trend component is highly significant, which can visually be assessed by the levels plots. Figure 14 - Figure 17 show the plots of the levels for demeaned data as well as their returns of the best ARMA(p,q)-GARCH(1,1). Additionally, the linear trend is plotted, as given in the legend. Figure 18 - Figure 21 show the plots of the detrended series and likewise the plot of their returns of the best ARMA(p,q)-GARCH(1,1) process. Additionally, Figure 22 plots the returns of the best ARMA(p,q)-GARCH(1,1) processes for adjusted limits of the y-axis, in order to illustrate that there is still fluctuation around zero.

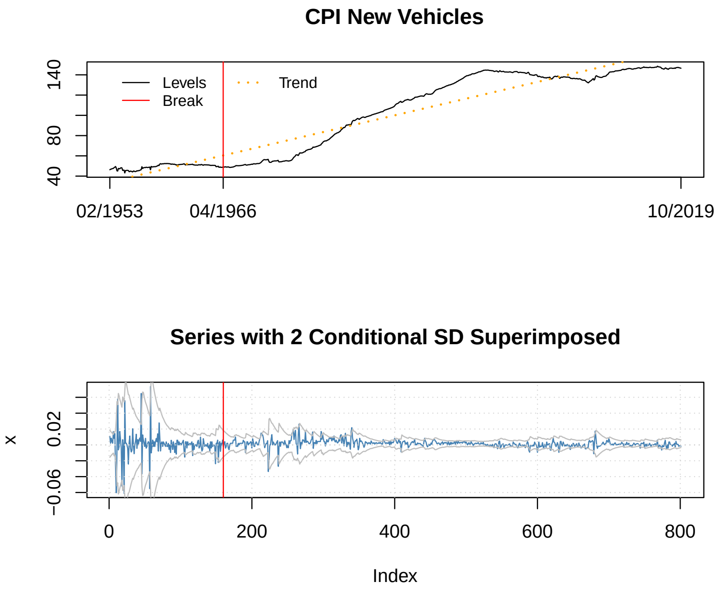

Figure 16: Plots of the demeaned time series and its GARCH fit.

4.5.2 Testing results

Table 17 and Table 18 summarize the estimated model fits, providing information on the significance of the estimated parameters, the goodness of fit and the testing results. For the bootstrapped Leybourne test, the p-value, p^b, is given along with the minimum and the maximum of the 999 bootstrapped test statistics in column $\widehat{ts}^b$. Analogously to the L test, bL rejects on the lower bound, whenever significantly many bootstrapped test statistics fall below the reference value, which is here the original L statistic. It is then concluded that the minimum of statistics of the forwarded series, see (8), is the breakpoint. Respectively, the breakpoint is given as the minimum of statistics of the reversed series, see (9), whenever the bootstrapped test statistics exceed the L test statistic, i.e.

$$ts^b < \widehat{ts} \Rightarrow BP = min(k^f)$$

$$ts^b > \widehat{ts} \Rightarrow BP = min(k^r).$$

Remember, that p^b is twice the relative amount of the minimum of the number of bootstrapped statistics that fall either below $\widehat{ts}$ or exceed it, i.e.
$p^b = 2\min\left(\frac{1}{B}\sum_{j=1}^{B}\mathbb{1}\left(ts_j^b \leq \widehat{ts}\right); \frac{1}{B}\sum_{j=1}^{B}\mathbb{1}\left(ts_j^b > \widehat{ts}.\right)\right)$, see (14).
For the interpretation of the findings in the empirical application, recall the following: Under the null hypothesis a constant instationary process is assumed (H_1) and tested against the alternative of a change in persistence, either from trend stationarity to difference stationarity, i.e. H_{01}, or vice versa, i.e. H_{10}. Hereby, the direction of change is not investigated and, hence, not determined. Pointing out a main in advance, only the bL finds breaks in persistence, however, only for the case of demeaned data. One remark on the estimated GARCH models: The parameterizations have illustrative character and serve rather as additional information on the goodness of the model fit.

In the case of *demeaned data*, Table 17 lists the model fits and results for the persistence

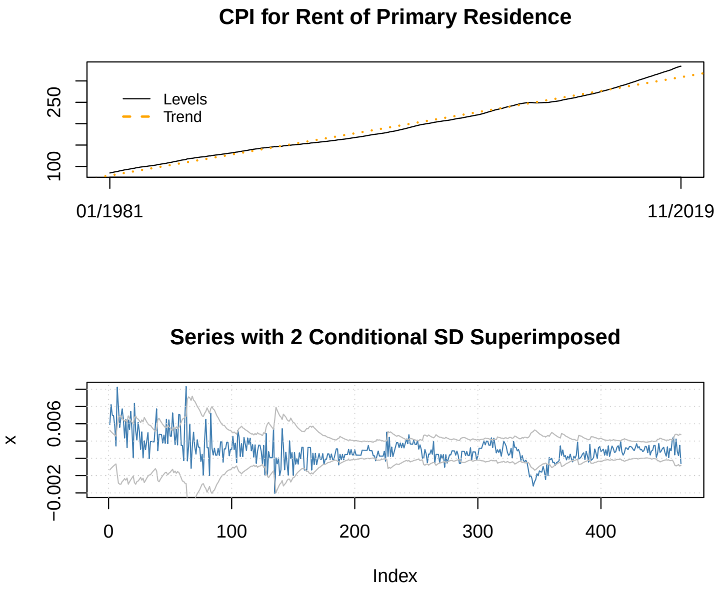

Figure 17: Plots of the demeaned time series and its GARCH fit.

change tests. It is noteworthy, that only the two longest series, i.e. the S&P and CPI Newvec, find a break in persistence. Nevertheless, CPI Newvec finds even the same breakpoint for both processes, the best parametrization ARMA(p,q)-GARCH(1,1) and the predefined AR(1)-GARCH(1,1), while the S&P finds a break only for the best parametrization. CPI Newvec finds the breakpoint in April 1966 at observation 160 and the S&P on the 08/26/1988 at observation 1966.

Regarding the breakpoint in CPI Newvec, c.f. the red vertical line in Figure 16, no obvious shock can be observed around that point in time, whereby one should bear in mind, that a CPI and not a stock index is considered. Referring to the plot of the levels and the plot of the fitted returns, respectively, changes in the behaviour of the series are observable, which may have caused the finding. The plot of the returns shows that the break occurs just after a period of higher volatility when the series changes to a less volatile *regime* or, at least, to a period that is less prone to many sequential peaks. Regarding the levels of the CPI Newvec, the breakpoint seems to induce a period of increases, which may be a possible persistence change.

Hence, it leaves the question, why the Leybourne test does not find a break. A possible explanation may be, that the time series is in fact stationary and, thus, the Leybourne test does not spuriously reject in the presence of the *false* null hypothesis, i.e. H_0.

Considering the break in the S&P, partly the same interpretation is applicable. The returns plot in Figure 14 seems to enter a phase of less volatility after a cluster with huge peaks has taken place, while in the levels a period of growth is introduced. Again, the Leybourne test may have prevented from spurious rejection in the face of the wrong initial situation, i.e. H_0.

It has to be stated, that in regard of the significance of parametrization, the GARCH parameters in the case of CPI Newvec and S&P are nearly all highly significant, while the GARCH parametrization of Consumption lacks high significance. As a matter of fact, this series may simply be too short in order to detect a persistence change.

In regard of the test statistics of bL and L, respectively, it is apparent that in the face of the less effective parametrization, i.e. the presumed AR(1)-GARCH(1,1), L experiences a distinct

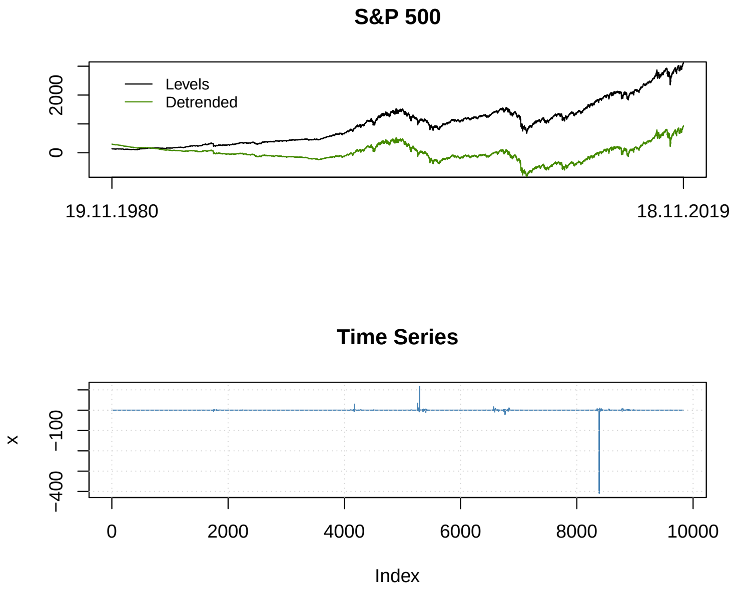

Figure 18: Plots of the detrended time series and its GARCH fit.

increase, e.g. from 0.88 to 1.13 (CPI Rent) or 0.62 to 1.40 (S&P) but not in case of Consumption, which is only 0.96 to 0.97. Again, this may be referred to the shortness of the series.

The distributions of the bootstrapped test statistics also shift to the right, but not as much as L. This is why it is even more difficult to find a break in persistence in the face of a less effective model specification, which is likewise underlined by the lower AICs, respectively. The AICs of the S&P are slightly distinct, as the actual less efficient parametrization, i. e. AR(1)-GARCH(1,1) has in fact the same AIC as the ARMA(3,1)-GARCH(1,1), which is probably caused by the length of the series.

For the bounds of the bootstrapped test statistics, a rule which explains the behavior cannot distinctly be derived. In the face of the seemingly less efficient parametrization, the extrema become narrower in case of CPI Rent and CPI Newvec, i.e. $[0.62; 1.39] \rightarrow [0.72; 1.33]$ and $[0.71, 1.53] \rightarrow [0.72, 1.50]$. In case of the S&P, the bounds shift to the right, i.e. from $[0.72; 1.61] \rightarrow [0.95; 1.88]$. A potential explanation is that the financial stock index underlies in fact different interdependencies than the CPIs, which is why the behaviour of the test statistics of *bl* and L are distinct as well as the finding of breakpoints.

Figure 22 depicts the returns of the series of *detrended data*. It is obvious that the returns exhibit enormous single peaks which is significantally different relative to the returns of the demeaned data. In levels, one peak has a value of -400, cf. Figure 18. This might indicate that the detrended data are not as well specified with ARMA(p,q)-GARCH(1,1) as the demeaned data. The finding is not directly supported by the significance of the parameters, but may hinted at by the higher AICs. The boundaries of the bL interval have become more extreme, cf. CPI Rent and CPI Newvec, respectively $[0.37; 1.48] \rightarrow [0.85; 1.33]$ and $[0.45, 2.01] \rightarrow [0.72, 1.01]$. Again, the behaviour of the bootstrap distribution of the S&P differs from the behavior of the other series. It is noteworthy that all test statistics of L peak around 1 and remain unaffected for the best ARMA(p,q)-GARCH(1,1) fit and the presumed AR(1)-GARCH(1,1).

Here, attention has to be drawn on the sum of GARCH parameters, which often indicate explosive behaviour. If the sum of $\alpha_1 + \beta$ is greater than one, the model is unstable. This is

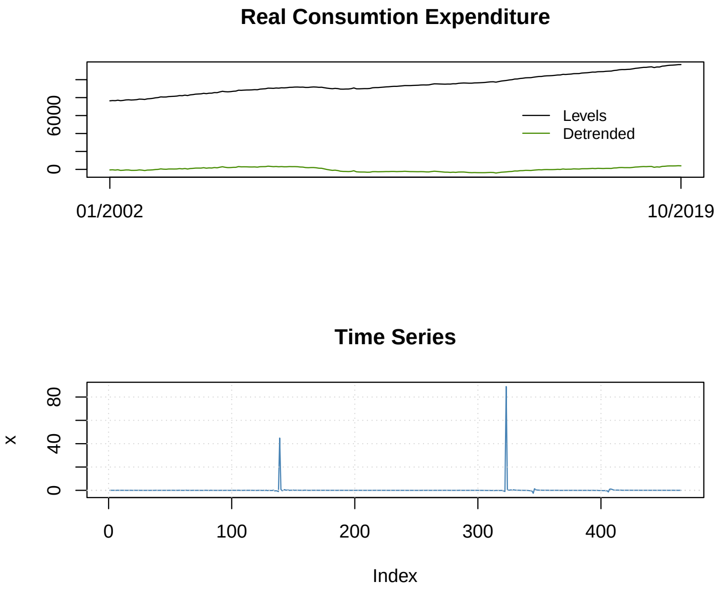

Figure 19: Plots of the detrended time series and its GARCH fit.

an indication that a stationary GARCH(1,1) model may not be adequate to fit the detrended data well - moreover, some estimators for the GARCH constant could not be computed due to singularity of the Hessian which also indicates explosive behaviour.

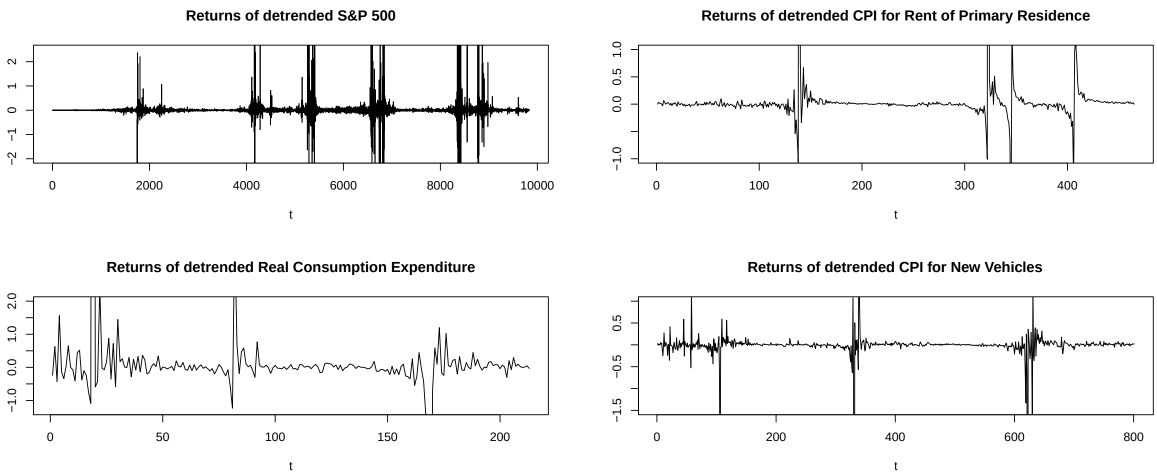

Figure 22: plots the returns of the four detrended time series.

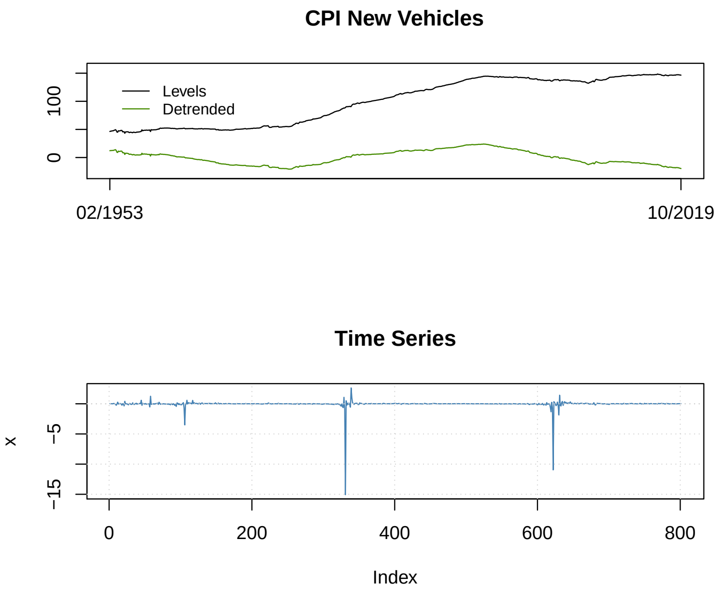

Figure 20: Plots of the detrended time series and its GARCH fit.

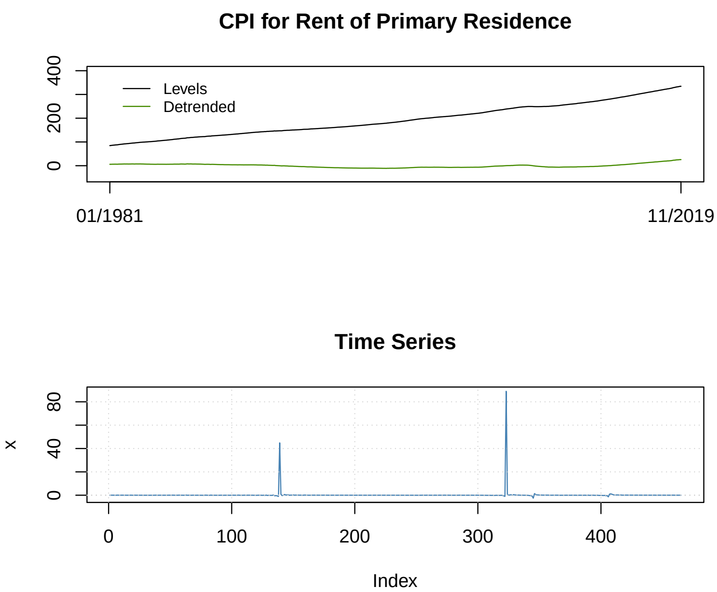

Figure 21: Plots of the detrended time series and its GARCH fit.

Table 17: Fitted GARCH-models for **demeaned data**, including information on: significance for trend & fit, goodness of fit, rejection behaviour of bL and L as well as the breakpoint.

MODEL SPECIFICATION		**Trend**	**AR**					**MA**					**GARCH**			**G.O.F.**		**bL**			**L**		
		levels	ϕ					θ					α_0	α_1	β	AIC	BIC	***rej.***	ts^b	**BP**	***rej.***	$\widehat{ts}$	**BP**
CONSUMPTION 214 data																							
ARMA(2,2)-GARCH(1,1)	Est	16.83	0.37	0.61				-0.39	-0.45				1.2e-06	0.15	0.73	-8.62	-8.52	p^b	[0.77	-	-	0.96	-
	SE	0.11	0.25	0.28				0.30	0.30				6.6e-07	0.07				=	;				
	sig	***		*				.					.	*	***			0.32	1.35]				
AR(1)-GARCH(1,1)	Est	16.83	0.27										1.1e-06	0.12	0.80	-8.34	-8.27	p^b	[0.79	-	-	0.97	-
	SE	0.11	0.07										8.0e-07	0.06	0.10			=	;				
	sig	***	***											.	***			0.37	1.39]				
PRIMARY RENT 466 data																							
ARMA(3,5)-GARCH(1,1)	Est	0.49	0.28	-0.08	0.80			-0.11	0.38	-0.72	0.20	-0.12	2.e-08	0.14	0.85	-11.14	-11.04	p^b	[0.62		-	0.88	-
	SE	2.6e-03	0.06	0.07	0.05			0.08	0.09	0.04	0.05	0.07	8.8e-09	0.03	0.03			=	;				
	sig	***	***		***				***	***	***	.	*	***	***			0.24	1.39				
AR(1)-GARCH(1,1)	Est	0.49	0.98										7.2e-08	0.4	0.57	-10.69	-10.65	p^b	[0.72	-	-	1.13	-
	SE	2.6e-03	0.01										2.7e-08	0.10	0.07			=	;				
	sig	***	***										**	***	***			0.16	1.33]				
NEW VEHICLES 802 data																							
ARMA(5,5)-GARCH(1,1)	Est	0.16	1.00	-0.50	0.90	-0.01	-0.40	-0.79	0.23	-0.68	-0.25	0.53	1.3e-07	0.16	0.86	-8.01	-7.93	p^b	[0.71	160	-	0.33	-
	SE	1.8e-03	0.15	0.16	0.13	0.16	0.15	0.14	0.12	0.10	0.12	0.12	4.8e-08	0.02	0.01			=	;				
	sig	***	***	*	***		**	***	.	***	*	***	**	***	***			0.00	1.53]				
AR(1)-GARCH(1,1)	Est	0.16	0.37										1.3e-07	0.14	0.87	-7.87	-7.85	p^b	[0.72	160	-	0.53	-
	SE	1.8e-03	0.04										5.7e-08	0.02	0.01			=	;				
	sig	***	***										*	***	***			0.00	1.50]				
S&P 500 9835 data																							
ARMA(3,1)-GARCH(1,1)	Est	0.24	0.71	-0.01	-0.02			-0.70					1.6e-06	0.09	0.90	-6.56	-6.56	p^b	[0.72	1966	-	0.62	-
	SE	9.8-04	0.10	0.01	0.01			0.10					2.1e-07	0.01	0.01			=	;				
	sig	***	***					***					***	***	***			0.00	1.61]				
AR(1)-GARCH(1,1)	Est	0.24	0.004										1.6e-06	0.09	0.90	-6.56	-6.55	p^b	[0.95	-	-	1.40	-
	SE	9.8e-04	0.01										2.1e-07	0.01	0.01			=	;				
	sig	***											***	***	***			0.56	1.88]				

Significance codes: $***$ = 0.000 ; $**$ = 0.001 ; $*$ = 0.01 ; . = 0.05 ; white space implies it is insignificant ($p \geq 0.1$). p^b denotes the bootstrapped p-value and by ts^b the minimum and maximum of the 999 bootstrapped test statistics are given. The auxiliary residuals are normally distributed.

Table 18: Table with fitted GARCH-models for **detrended data**, including information on: significance for fit, goodness of fit, rejection behaviour of bL and L as well as the breakpoint.

MODEL SPECIFICATION		**AR**		**MA**		**GARCH**			**G.O.F.**		**bL**			**L**		
		ϕ		θ		α_0	α_1	β	AIC	BIC	***rej.***	ts^b	**BP**	***rej.***	$\widehat{ts}$	**BP**
CONSUMPTION 214 data																
	Est	-0.95		0.95		9.1e-03	1.00	0.33	0.20	0.32	p^b	[0.61	-	-	1.00	-
ARMA(1,1)-GARCH(1,1)	SE	0.03		0.03		NA	0.20	0.10			=	;				
	sig	**		***			***	**			0.52	2.80]				
	Est	-3.4e-03				0.01	1.00	0.36	0.20	0.30	p^b	[0.79	-	-	1.19	-
AR(1)-GARCH(1,1)	SE	0.07				NA	0.3	0.10			=	;				
	sig						**	***			0.53	2.50]				
PRIMARY RENT 466 data																
	Est	0.89		-0.52		4.0e-04	1.00	0.21	-3.17	-3.11	p^b	[0.37	-	-	0.99	-
ARMA(1,1)-GARCH(1,1)	SE	0.02		0.05		2.1e-03	0.22	0.04			=	;				
	sig	***		***			***	***			0.61	1.48]				
	Est	0.66				8.9e-04	1.00	0.07	-2.99	-2.94	p^b	[0.85	-	-	1.00	-
AR(1)-GARCH(1,1)	SE	0.05				1.6e-03	0.19	0.05			=	;				
	sig	***					***				0.99	1.13]				
NEW VEHICLES 802 data																
	Est	0.69	0.26	-0.16	-0.58	1.0e-04	1.00	0.53	-2.51	-2.47	p^b	[0.45	-	-	1.00	-
ARMA(2,2)-GARCH(1,1)	SE	0.05	0.04	0.03	0.02	NA	0.09	0.02			=	;				
	sig	***	***	***	***		***	***			0.94	2.01]				
	Est	0.52				8.2e-05	0.89	0.60	-2.43	-2.41	p^b	[0.72	-	-	1.00	-
AR(1)-GARCH(1,1)	SE	0.04				5.7e-08	0.01	0.01			=	;				
	sig	***					***	***			0.98	1.11]				
S&P 500 9835 data																
	Est	-0.26		0.26		1.8e-05	0.56	0.07	-3.09	-3.08	p^b	[0.46	-	-	1.00	-
ARMA(1,1)-GARCH(1,1)	SE	0.78		0.78		2.6e-04	0.04	8.3e-03			=	;				
	sig						***	***			0.97	2.10]				
	Est	1.7e-03				1.9e-05	0.56	0.70	-3.08	-3.09	p^b	[0.44	-	-	1.00	-
AR(1)-GARCH(1,1)	SE	0.01				3.4e-04	0.06	8.6e-03			=	;				
	sig						***	***			0.98	2.11]				

Significance codes: $*\!*\!* = 0.000$; $** = 0.001$; $* = 0.01$; $. = 0.05$; white space implies it is insignificant ($p \geq 0.1$). p^b denotes the bootstrapped p-value and by ts^b the minimum and maximum of the 999 bootstrapped test statistics are given. The auxiliary residuals are normally distributed.

4.6 Conclusion

In this work it is investigated whether structural changes in the conditional volatility may render the persistence change test proposed by Leybourne et al. (2007) to falsely reject the null hypothesis of no persistence change. For robustification purposes the Leybourne test is bootstrapped and three different auxiliary distributions are employed. Within a simulation study, power and size results are derived for various GARCH processes and compared to its original counterpart. The main finding is that the bootstrapped test is heavily oversized as soon as ϕ increases. Due to the oversizement the power results are questionable. The L test seems to meet the size very well in most cases, although it is very conservative in the face of the *false* null hypothesis, meaning constantly stationary behavior. Nevertheless, explosive behavior leads to extremely low power results in case of a switch from H_{10}. It is pointed out that the Leybourne test seems heavily impacted when there are additional changes in the conditional volatility in especially very finite sample sizes, what justifies the purpose of this investigation. Nevertheless, the proposed bootstrapped version of the Leybourne test is incapable to alleviate size distortions or to improve the testing outcomes of the L test.

Within an empirical application both versions of the Leybourne et al. (2007) test are applied to three inflation rates and one stock market index. The finding is that the bootstrapped test succeeds to detect breaks in the demeaned but not in the detrended data, whereas the original Leybourne et al. (2007) test never rejects in favour of a break in persistence. This may be referred to the fact that the Leybourne test is very conservative in the face of the wrong null hypothesis as suggested by the authors. The Leybourne test does not spuriously reject in favour of a persistence change, which can be concluded from both the simulation study and the empirical application. Different conclusions have to be drawn for very short time series horizons, e.g. for the Consumption inflation rate. As a result, it is questionable whether the presumed AR(1)-GARCH(1,1) or other ARMA(p,q)-GARCH(1,1) models are suitable to model the properties of the detrended time series. Furthermore, the results of the empirical investigation suggest that inflation rates behave differently than stock market indices. However, it has not been clarified yet in how far the additional nuisance parameters affect the limiting distribution of the tests, which is left for future investigations.

Bibliography

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

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

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

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

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

Antoch, J., Hanousek, J., Horvath, L., Huankova¡, M., and Wang, S. (2019). Structural breaks in panel data: Large number of panels and short length time series. *Econometric Reviews*, 38:828–855.

Beran, R. (1986). Discussion: Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14(4):1295–1298.

Bierens, H. and Martins, L. (2010). Time-varying cointegration. *Econometric Theory*, 26:1453—-1490.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

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

Busetti, F. and Taylor, A. (2003). Variance shifts, structural breaks, and stationarity tests. *Journal of Business & Economic Statistics*, 21:510–531.

Busetti, F. and Taylor, A. (2004). Tests of stationarity against a change in persistence. *Journal of Econometrics*, 123:33–66.

Cavaliere, G. (2004). Testing stationarity under a permanent variance shift. *Economics Letters*, 82:403–408.

Cavaliere, G. (2005). Unit root tests under time-varying variances. *Econometric Reviews*, 23:259–292.

Cavaliere, G. and Taylor, A. (2006). Testing for a change in persistence in the presence of non-stationary volatility. Discussion Papers 06/04, University of Nottingham, Granger Centre for Time Series Econometrics.

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

Cavaliere, G. and Taylor, A. (2008). Testing for a change in persistence in the presence of non-stationary volatility. *Journal of Econometrics*, 147:84–98.

Clements, M. and Hendry, D. (1996). Intercept corrections and structural change. *Journal of Applied Econometrics*, 11:475–494.

Davidson, J., Hendry, D., Srba, F., and Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the united kingdom. *The Economic Journal*, 88:661–669.

Davidson, J., Monticini, A., and Peel, D. (2007). Implementing the wild bootstrap using a two-point distribution. *Economics Letters*, 96:309–315.

Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, 146:162–169.

Davidson, R. and MacKinnon, J. G. (2000). Bootstrap tests: How many bootstraps? *Econometric Reviews*, 19(1):55–68.

DeJong, D., Nankervis, J., Savin, N., and Whiteman, C. (1992a). Integration versus trend stationary in time series. *Econometrica*, pages 423–433.

DeJong, D., Nankervis, J., Savin, N., and Whiteman, C. (1992b). The power problems of unit root test in time series with autoregressive errors. *Journal of Econometrics*, 53:323–343.

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

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

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1007.

Engle, R. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5:1–50.

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

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

Granger, C. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16:121–130.

Haldrup, N. and Jansson, M. (2006). *Palgrave Handbook of Econometrics 1*, chapter Improving size an power in the unit root testing, pages 252–277. Palgrave Macmillan.

Hansen, B. (1991). Testing for structural change of unknown form in models with non-stationary regressors.

Hansen, B. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, pages 413–430.

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

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

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

Hsieh, D. (1995). Nonlinear dynamics in financial markets: Evidence and implications. *Financial Analysts Journal*, 51:55–62.

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

Jockel, K.-H. (1986). Finite sample properties and asymptotic efficiency of monte carlo tests. *The annals of Statistics*, 14(1):336–347.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59:1551–1580.

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

Kapetanios, G., Shin, Y., and Snell, A. (2006). Testing for cointegration in nonlinear smooth transition error correction models. *Econometric Theory*, 22:279—-303.

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

Kiliç, R. (2011). Testing for co-integration and nonlinear adjustment in a smooth transition error correction model. *Journal of Time Series Analysis*, 32:647—660.

Kim, J. (2000). Detection of change in persistence of a linear time series. *Journal of Econometrics*, 95:97–116.

Kim, J., Belaire, F., and Amador, R. (2002). Corrigendum to "detection of change in persistence of a linear time series"[j. econom. 95 (2000) 97–116]. *Journal of Econometrics*, 109:389–392.

Kwiatkowski, D., Phillips, P., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54:159–178.

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

Leybourne, S., Kim, T.-H., Smith, V., and Newbold, P. (2003). Tests for a change in persistence against the null of difference-stationarity. *The Econometrics Journal*, 6:291–311.

Leybourne, S. and Taylor, A. (2004). On tests for changes in persistence. *Economics Letters*, 84:107–115.

Leybourne, S., Taylor, A., and Kim, T.-H. (2007). CUSUM of squares-based tests for a change in persistence. *Journal of Time Series Analysis*, 28:408–433.

Liu, R. (1988). Bootstrap procedures under some non-iid models. *The Annals of Statistics*, 16(4):1696–1708.

MacKinnon, J. G. (2009). Bootstrap hypothesis testing. *Handbook of computational econometrics*, 183:213.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, pages 255–285.

McCabe, B. and Harrison, M. (1980). Testing the constancy of regression relationships over time using least squares residuals. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29:142–148.

Nyblom, J. and Mäkeläinen, T. (1983). Comparisons of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association*, 78:856–864.

Pearson, K. (1916). Mathematical contributions to the theory of evolution. xix. second supplement to a memoir on skew variation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, pages 429–457.

Ploberger, W. and Krämer, W. (1992). The CUSUM test with ols residuals. *Econometrica*, 60(2):271–285.

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

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

Taylor, A. (2005). Fluctuation tests for a change in persistence. *Oxford Bulletin of Economics and Statistics*, 67:207–230.

Taylor, D. H. S. L. A. (2006). Modified tests for a change in persistence. *Journal of Econometrics*, 134:441–469.

Taylor, M., Peel, D., and Sarno, L. (2001). Nonlinear mean reversion in real exchange rates: Toward a solution to the purchasing power parity puzzles. *International Economic Review*, 42:1015—-1042.

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

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

White, H. et al. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838.

Wu, C.-F. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, pages 1261–1295.

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

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

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

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

Zivot, E. and Andrews, D. (2002). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics*, 20:25–44.