Change–point detection in GARCH models: asymptotic and bootstrap tests

Piotr Kokoszka
Mathematics and Statistics
Utah State University

Gilles Teyssière
National Bank of Greece, France
GREQAM & CORE

November 25, 2002

Abstract

Two classes of tests designed to detect changes in volatility are proposed. Procedures based on squared model residuals and on the likelihood ratio are considered. The tests are applicable to parametric nonlinear models like GARCH. Both asymptotic and bootstrap tests are investigated by means of a simulation study and applied to returns data. The tests based on the likelihood ratio are shown to be generally preferable. A wavelet based estimator of long memory is applied to returns data to shed light on the interplay of change points and long memory.

Key words: GARCH model, change–point, likelihood ratio, parametric bootstrap, squared residuals, size–power curves, wavelets.

Running head: Change–point in GARCH

1 Introduction

The objective of this paper is to propose and investigate by means of simulations and applications to return data several tests for detecting a change–point in a time series which can be approximated by a GARCH model. Such series are often obtained as residuals of linear time series models like ARIMA or classical decomposition models (trend, seasonal component, stationary component).

A very large body of empirical and theoretical research on detecting parameter changes in independent or linearly dependent observations has accumulated over the last two decades. It is not possible to equitably give credit to major contributions in this area, so we merely mention the special volume of *Journal of Business and Economic Statistics* on Breakpoints and Unit Roots, July 1992, and the monograph of Csörgő and Horváth (1997). Some further references can also be found in Andreou and Ghysels (2001) who however focus on detecting changes in volatility which exhibits *nonlinear* dependence. The goal of the present paper is similar, but unlike Andreou and Ghysels (2001), who considered nonparametric tests motivated by the works of Kokoszka and Leipus (1998, 1999, 2000) and Lavielle and Moulines (2000), we investigate parametric tests which could *a priori* work in situations where the parameters change but the unconditional variance does not change or changes very little.

We examine two types of change–point tests: tests based on the sequential empirical process of the squared residuals and generalized likelihood ratio (GLR) tests. The theory underlying the former tests was developed by Horváth *et al.* (2001) and Berkes and Horváth (2002). No theory is at present available for the GLR tests in the GARCH context, but given the importance of such procedures in change–point detection methodology, see e.g. Csörgő and Horváth (1997) and Besseville and Nikifirov (1993), we think it is important to assess their performance by simulation. We hope that this will motivate both theoretical research in this direction and the application of these tests in more complex settings, e.g. like those considered by Andreou and Ghysels (2001) and Kirman and Teyssière (2002a, 2002b). We present our findings using traditional tabulation and modern graphical methods like size-power curves developed by Davidson and MacKinnon (1998) and used by Mantalos (2000) and Horváth *et al.* (2002), among others.

There is small but growing literature on detecting change–points in volatility. In addition to the aforementioned work, we mention Chu (1995) and Lundbergh and Teräsvirta (2002) who consider Lagrange multiplier tests and Mikosch and Stărică (1999, 2002a) who developed tests based on the periodogram of the data, but did not systematically investigate their empirical performance. Berkes *et al.* (2002) developed on-line change–point detection tests for GARCH which are based on likelihood scores. Related research is also presented in Horváth *et al.* (2002), but it is mainly concerned with model misspecification tests.

Increasing attention has also been given to the effect of possible change–points on tests designed to detect long memory and vice versa. This interesting controversy underlines the need for further research on robust statistical procedures for both long memory and change–point models, see e.g. Diebold and Inoue (2001), Giraitis *et al.* (2001), Mikosch and Stărică (2002a) and Kokoszka and Leipus (2002).

The detection of change–points in volatility processes is also of interest for macroeconomic analysis. Mikosch and Stărică (1999) observed that the detected changes in the unconditional variance implied by changes in the parameters of a fitted GARCH(1,1) model match the transitions from economic expansion to recession and vice versa. A change in the conditional volatility of macroeconomic business–cycle series has been also noticed by Stock and Watson (2002).

The paper is organized as follows. In Section 2 we describe tests based on the sequential empirical process of squared residuals, Section 3 is devoted to the GLR tests. Section 4 contains the details of the simulation study. Section 5 reports the application of these tests to series of daily returns on financial assets.

We use the classical GARCH(p, q) model to describe and compare the various tests, but it is clear that our procedures can be adapted in a straightforward manner to any parametric model for which residuals can be computed or a likelihood function written down.

Recall that observations $y_1, \dots, y_n$ follow a GARCH(p, q) model if they satisfy the equations:

$$y_k = \sigma_k \varepsilon_k, \tag{1.1}$$

$$\sigma_k^2 = \omega + \sum_{j=1}^{p} \beta_j \sigma_{k-j}^2 + \sum_{j=1}^{q} \alpha_j y_{k-j}^2. \tag{1.2}$$

The *innovations* ε_k in (1.1) are independent and identically distributed and $\omega, \alpha_j, \beta_j$ are nonnegative parameters. The parsimonious GARCH$(1, 1)$ model is often used for exploratory analysis because it is simple (only three parameters to be estimated) and captures the dependence in volatility to a degree that allows to conduct further useful inference, see e.g. Chapter 15 of Hull (2000).

2 Change–point detection using the sequential empirical process of squared residuals

Suppose we observe a sample $y_1, \dots, y_n$ which we model by a GARCH(p, q) process. The method described in this subsection is *a priori* most suitable for detecting a change in the distribution of the unobservable innovations ε_k but can be expected to detect parameter changes as well.

After estimating the parameters of the model, we can construct estimated conditional variances $\hat{\sigma}_i^2$ by replacing in (1.2) the parameters by their estimates. We can then construct the residuals $\hat{\varepsilon}_i^2 = y_i^2 / \hat{\sigma}_i^2$ and their empirical process

$$\hat{e}_n(t) = n^{-1/2} \sum_{1 \le i \le n} \left[\mathbf{I} \left\{ \hat{\varepsilon}_i^2 \le t \right\} - F(t) \right], \tag{2.1}$$

where F is the distribution function of ε_i^2 (e.g. $F \sim \chi_1^2$ if the ε_i are standard normal). In Horváth *et al.* (2002) we examined bootstrap goodness-of-fit tests based on (2.1). These tests were intended to check if the observed time series can be reasonably approximated by a GARCH model of a specified order. We expected that they will also detect a change in the distribution of the innovations ε_i^2, but simulations actually showed that this is not the case: typically a test of size 5% will have power of only about 10% against a fairly large change in the distribution of the ε_i^2. For other departures from the GARCH specification this power may approach 100%.

We therefore focus here on tests based on the *sequential* empirical process:

$$\hat{K}_n(s,t) = n^{-1/2} \sum_{1 \le i \le [ns]} \left[\mathbf{I} \left\{ \hat{\varepsilon}_i^2 \le t \right\} - F(t) \right], \quad 0 < s \le 1. \tag{2.2}$$

As s increases from 0 to 1, this process "looks" at all potential change–points. More specifically, the proposed tests are based on the process

$$\hat{T}_n(s,t) = \sqrt{n} \frac{[ns]}{n} \left(1 - \frac{[ns]}{n} \right) \left(\hat{F}_{[ns]}(t) - \hat{F}^*_{n-[ns]}(t) \right), \tag{2.3}$$

where

$$\hat{F}_{[ns]}(t) = \frac{1}{[ns]} \sum_{1 \le i \le [ns]} \mathbf{I} \left\{ \hat{\varepsilon}_i^2 \le t \right\} \tag{2.4}$$

and $F^*_{n-[ns]}(t)$ is defined analogously using the residuals with indexes larger than $[ns]$.

A process analogous to (2.3) was used by Bai (1994) to detect changes in the distribution of the innovations in ARMA models. In that case, process (2.2) converges to the Kiefer process $K(s,t)$, see e.g. Shorack and Wellner (1986) for the definition of $K(s,t)$ (if s is fixed this process is a Brownian bridge and if t is fixed it is a Brownian motion). Horváth *et al.* (2001) showed that this is not true for ARCH(p). In fact, for ARCH(p)

$$\hat{K}_n(s,t) \xrightarrow{d} K(s,t) + stf(t)\xi, \tag{2.5}$$

where f is the density of ε_i^2 and ξ is a Gaussian random variable correlated with $K(s,t)$ in a complex way. Observe however that if (2.5) holds, then (see Example 3.1 in Horváth *et al.* (2001))

$$\hat{T}_n(s,t) = \hat{K}_n(s,t) - \frac{[ns]}{n} \hat{K}_n(1,t) \tag{2.6}$$

$$\sim (K(s,t) + stf(t)\xi) - s\,(K_n(1,t) + tf(t)\xi)$$

$$= K(s,t) - sK(1,t), \tag{2.7}$$

so, unlike for (2.1), the troublesome term $stf(t)\xi$ disappears. Process (2.7) is the Gaussian process of Bickel and Wichura (1971) and has been extensively studied. For example its supremum is tabulated in Picard (1985). Thus a useful test statistic could be $\sup_{0 \le s \le 1} \sup_{t > 0} \left| \hat{T}_n(s,t) \right|$. Other functionals can also be considered, as we shall see in the sequel.

Berkes and Horváth (2002) showed that (2.5)–(2.7) hold not only for ARCH(p) but also for general GARCH(p,q) models. Their results effectively only require that the innovations and observations have a finite moment of an order arbitrarily close to zero.

We focus on two specific test statistics based on the process (2.3). Both of them are functionals of the absolute value of the process (2.3) with known asymptotic distribution. The idea is that if there is a change in the model, the absolute value of $\hat{T}_n(s,t)$ is "large" for some s and t.

The Kolmogorov-Smirnov type statistic: For $1 \le k \le n$ define

$$\hat{F}_k(t) = \frac{1}{k}\#\{i \le k : \hat{\varepsilon}_i^2 \le t\}, \qquad \hat{F}_k^*(t) = \frac{1}{n-k}\#\{i > k : \hat{\varepsilon}_i^2 \le t\}; \tag{2.8}$$

$$\hat{T}_n(k,t) = \sqrt{n} \cdot \frac{k}{n}\left(1 - \frac{k}{n}\right)\left|\hat{F}_k(t) - \hat{F}_k^*(t)\right|; \tag{2.9}$$

$$\hat{M}_n = \sup_{0 \le t \le \infty} \max_{1 \le k \le n} |\hat{T}_n(k,t)| = \max_{1 \le k \le n} \max_{1 \le j \le n} |\, \hat{T}_n(k, \hat{\varepsilon}_j^2) \,| \,. \tag{2.10}$$

The asymptotic distribution of $\hat{M}_n$ is the same as for Picard's (1985) generalized Kolmogorov-Smirnov statistic. Table 1, which is based on Table 1 on p. 845 of Picard (1985), gives the asymptotic critical values c for the significance level α, i.e. $P(\hat{M}_n > c) \sim \alpha$.

α	c	α	c
0.10	0.715	0.02	0.823
0.05	0.772	0.01	0.874

Table 1: Asymptotic critical values for the Kolmogorov-Smirnov statistic (2.10): $P(\hat{M}_n > c) \sim \alpha$.

The Cramer-VonMises type statistic: Blum *et al.* (1961) tabulated the distribution function of the random variable

$$B = \int_0^1 \int_0^1 T^2(s,u)\,du\,ds, \tag{2.11}$$

where T is the "tied-down" Kiefer process, i.e. a Gaussian field on $[0,1] \times [0,1]$ with the covariance structure

$$\begin{aligned} E[T(s,u)] &= 0, \\ E[T(x,y)T(s,u)] &= (x \wedge s - xs)(y \wedge u - yu), \end{aligned}$$

where $a \wedge b = \min(a,b)$. Thus, $T(s,t)$ has the same distribution as the process (2.7).

Recall definition (2.9) and introduce the random field

$$\hat{W}_n(s,t) = \hat{T}_n([ns],t). \tag{2.12}$$

The process $\hat{W}_n(s,t)$ converges in an appropriate Skorokhod space to $T(s, F(t))$, where F is the cdf of the ε_i^2. It can be expected that the distribution of

$$\hat{B}_n = \int_0^1 \int_0^\infty \left[\hat{T}_n([ns],t)\right]^2 d\hat{F}(t)\,ds \tag{2.13}$$

will, for large n, be well approximated by the distribution of B. To give a rough justification to this claim observe that

$$
\begin{aligned}
\hat{B}_n &= \int_0^1 \left[\int_0^\infty [\hat{W}_n(s,t)] d\hat{F}(t)\right] ds \\
&\approx \int_0^1 \left[\int_0^\infty [T(s,F(t))]^2 dF(t)\right] ds \\
&= \int_0^1 \left[\int_0^1 [T(s,u)]^2 du\right] ds \qquad (u = F(t)) \\
&= B.
\end{aligned}
$$

The double integral defining $\hat{B}_n$ can be computed as a finite sum:

$$
\hat{B}_n = \int_0^1 \left\{ \frac{1}{n} \sum_{i=1}^n [\hat{T}_n([ns], \hat{\varepsilon}_i^2)]^2 \right\} ds \tag{2.14}
$$

$$
= \frac{1}{n^2} \sum_{k=1}^n \sum_{i=1}^n [\hat{T}_n(k, \hat{\varepsilon}_i^2)]^2. \tag{2.15}
$$

The cdf of $\frac{1}{2}\pi^4 B$ and its upper quantiles can be found on p. 497 of Blum *et al.* (1961). Table 2 is based on Table II of Blum *et al.* (1961).

α	$q(\alpha)$	α	$q(\alpha)$
0.1	0.04694	0.005	0.09960
0.05	0.05839	0.001	0.12976
0.01	0.08685	0.0005	0.14290

Table 2: Upper quantiles of B; $P(B > q(\alpha)) = \alpha$.

3 Change–point detection using the generalized likelihood ratio test

The generalized likelihood ratio (GLR) test is perhaps the most commonly used approach for testing composite hypotheses, see e.g. Chapter 9 of Rice (1995). In this section we derive the form of the maximally selected GLR statistic for change–point detection in GARCH processes. If the observations are iid, there is a profound theory underlying the use of this statistics, see Csörgő and Horváth (1997). As far as we know, no such theory has yet been developed for GARCH processes.

In order to avoid cumbersome notation we present our derivation for the ARCH(1) model, the general case being similar.

Under the null hypothesis of no change in the data, the observations y_t follow the model

$$
y_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \omega + \alpha y_{t-1}^2, \qquad \varepsilon_t \sim \text{iid } N(0,1), \qquad t = 1, \ldots, n, \tag{3.1}
$$

The alternative is that the parameter vector (ω, α) changed at some unknown time t_0. Suppose for a moment that we *know* t_0. Thus for $t > t_0$ the observations follow the model

$$y_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \omega^* + \alpha^* y_{t-1}^2, \qquad \varepsilon_t \sim \text{iid } N(0,1),$$

where $\omega^* \neq \omega$ or $\alpha^* \neq \alpha$. In this case the GLR statistic is defined by

$$\Lambda_{t_0} = \frac{\text{maximum likelihood under null}}{\text{maximum likelihood if change at } t_0}. \tag{3.2}$$

The numerator in (3.2) is

$$\left(\prod_{t=1}^{n} [2\pi(\hat{\omega} + \hat{\alpha} y_{t-1}^2)]^{-\frac{1}{2}} \right) \exp\left\{ -\frac{1}{2} \sum_{t=1}^{n} \frac{y_t^2}{\hat{\omega} + \hat{\alpha} y_{t-1}^2} \right\},$$

where $\hat{\omega}$ and $\hat{\alpha}$ are the MLE's, and on setting $\hat{\omega} + \hat{\alpha} y_{t-1}^2 = \hat{\sigma}_t^2$, it becomes

$$(2\pi)^{-\frac{n}{2}} \left(\prod_{t=1}^{n} \hat{\sigma}_t^2 \right)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{n} \frac{y_t^2}{\hat{\sigma}_t^2} \right\}. \tag{3.3}$$

Let $\tilde{\omega}$, $\tilde{\alpha}$, $\tilde{\sigma}$, be the estimates based on $y_1, \ldots, y_{t_0}$ and $\bar{\omega}$, $\bar{\alpha}$, $\bar{\sigma}$ the estimates based on $y_{t_0+1}, \ldots, y_n$. Defining $\bar{\sigma}$ and $\tilde{\sigma}$ analogously to $\hat{\sigma}$, we see that the denominator in (3.2) is

$$(2\pi)^{-\frac{n}{2}} \left(\prod_{t=1}^{t_0} \tilde{\sigma}_t^2 \right)^{-\frac{1}{2}} \left(\prod_{t=t_0+1}^{n} \bar{\sigma}_t^2 \right)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \Big[\sum_{t=1}^{t_0} \frac{y_t^2}{\tilde{\sigma}_t^2} + \sum_{t=t_0+1}^{n} \frac{y_t^2}{\bar{\sigma}_t^2} \Big] \right\}, \tag{3.4}$$

and so

$$-2\log\Lambda_{t_0} = -\left[\sum_{t=1}^{t_0} (\log\tilde{\sigma}_t^2 - \log\hat{\sigma}_t^2) + \sum_{t=t_0+1}^{n} (\log\bar{\sigma}_t^2 - \log\hat{\sigma}_t^2) \right] + \sum_{t=1}^{n} \frac{y_t^2}{\hat{\sigma}_t^2} - \sum_{t=1}^{t_0} \frac{y_t^2}{\tilde{\sigma}_t^2} - \sum_{t=t_0+1}^{n} \frac{y_t^2}{\bar{\sigma}_t^2}.$$

Since, in fact, we do not know that a change took place at t_0, we consider the statistic

$$\begin{aligned} \Lambda_n^* &= \max_{1 \le k < n} (-2\log\Lambda_k) \\ &= \max_{1 \le k < n} \left[\sum_{t=1}^{n} \Big(\frac{y_t^2}{\hat{\sigma}_t^2} + \log\hat{\sigma}_t^2 \Big) - \sum_{t=1}^{k} \Big(\frac{y_t^2}{\tilde{\sigma}_t^2} + \log\tilde{\sigma}_t^2 \Big) - \sum_{t=k+1}^{n} \Big(\frac{y_t^2}{\bar{\sigma}_t^2} + \log\bar{\sigma}_t^2 \Big) \right]. \end{aligned} \tag{3.5}$$

Formula (3.5) is valid for any GARCH model with appropriately defined $\hat{\sigma}, \bar{\sigma}, \tilde{\sigma}$. For GARCH$(p,q)$ the initial values $\hat{\sigma}_0, \ldots, \hat{\sigma}_{p-1}$ etc. must be chosen somewhat arbitrarily. As recommended in Bollerslev (1986) and McCullough and Vinod (1999) we set $\hat{\sigma}_i^2$ and y_i^2, $i \le 0$ equal to the mean of the squared residuals.

Even if the observations are iid, the statistic Λ_n^* exhibits a very interesting behavior: it does not converge to a limiting distribution for any normalizing sequence. Instead it satisfies an Erdős -Darling limit theorem with an exponential type extreme value distribution as the limit, see Csörgő and Horváth (1997) for a detailed exposition. It follows from Theorem 1.3.1 of Csörgő and Horváth (1997) that for iid observations with estimated d parameters, the asymptotic size α test rejects the null hypothesis of no change in parameters if $\Lambda_n^* > c_n(\alpha)$ where

$$c_n(\alpha) = \frac{[D_d(\log n) - \log[-\log(1-\alpha)] + \log 2]^2}{2\log\log n}, \tag{3.6}$$

and where

$$D_d(x) = 2\log x + \frac{d}{2}\log\log x - \log\Gamma\left(\frac{d}{2}\right).$$

The asymptotic behavior of Λ_n^* (3.5) for GARCH processes is at present unknown. If the residuals $\hat{\varepsilon}_i^2$ are sufficiently close to the squared innovations ε_i^2, then Λ_n^* exhibits the same limit behavior as for the iid observations and the critical value (3.6) can be used. It may be the case, however, that the limiting distribution explicitly depends on model parameters, as in a different contexts in Horváth *et al.* (2001) and Berkes *et al.* (2002). This remains an interesting theoretical problem which requires clarification.

In the following we consider an asymptotic test which rejects the null if $\Lambda_n^* > c_n(\alpha)$ and a bootstrap test constructed as follows:

1. Calculate from the data the observed value λ_n^* of the statistic (3.5).

2. Generate B GARCH series of length n each with the order and parameters estimated from the data, the number of bootstrap samples B being chosen using a pretesting procedure advocated by Davidson and MacKinnon (2000) which is described below.

3. Calculate the corresponding B bootstrap values of Λ_n^*.

4. Determine the bootstrap P–value $\hat{P}^*$ as the proportion of the B values Λ_n^* calculated in the previous step which are greater than λ_n^*:

$$\hat{P}^* = \frac{1}{B}\sum_{i=1}^{B} \mathbf{I}[\Lambda_{n_i}^* > \lambda_n^*]. \tag{3.7}$$

In the last step we calculate the empirical bootstrap P–value $\hat{P}^*$. When $B \to \infty$, $\hat{P}^* \to P^*$, the ideal bootstrap P–value, i.e. the probability that $\Lambda_{n_i}^* > \lambda_n^*$ under the bootstrap DGP. The quality of this approximation clearly depends on the number of bootstrap samples B which we choose using the following data driven procedure of Davidson and MacKinnon (2000): Let $\hat{\tau}$ be the statistic of interest:

1. Compute $\hat{\tau}$, set $B = B' = 99$, and the statistics τ_j^*, $j = 1, \ldots, B$.

2. Compute the estimated bootstrap P-value $\hat{P}^*(\hat{\tau})$ based on the B bootstrap samples. Depending whether $\hat{P}^*(\hat{\tau}) < \alpha$ or $\hat{P}^*(\hat{\tau}) > \alpha$, test, at the level $\beta = 0.001$, either the hypothesis that $P^*(\hat{\tau}) \geq \alpha$ or $P^*(\hat{\tau}) \leq \alpha$. If the hypothesis is rejected, then stop, else go to step 3. Since we consider different values for α, i.e. tests at sizes 10%, 5% , 1%, we slightly modify the testing procedure.

3. Set $B = 2B' + 1$. If B is too large, e.g. $B > B_{max} = 12,799$ then stop, else calculate τ_j^* for a further $B' + 1$ bootstrap samples, set $B' = B$ and return to step 2.

Because of the peculiar behavior of Λ_n^* even in the iid case, one often focuses on statistics based on the stochastic process $\left\{t(1-t)(-2\log\Lambda_{[nt]}),\ 0 < t < 1\right\}$, see Section 1.1 of Csörgő and Horváth (1997) for further details.

If the observations are independent with a density depending on d parameters, then the above process can be approximated by $\left\{\sum_{i=1}^{d} B_i^2(t),\ 0 < t < 1\right\}$, where $B_i(\cdot), i = 1, \ldots, d$ are independent Brownian bridges on $[0,1]$. It then follows, for example, that

$$\Delta_n^* := n^{-3}\sum_{k=1}^{n-1} k(n-k)(-2\log\Lambda_k) \xrightarrow{d} \int_0^1 \sum_{i=1}^{d} B_i^2(t)dt. \tag{3.8}$$

Other functionals can also be considered and many of them are tabulated, see Csörgő and Horváth (1997) p. 6. Using Table 4 in Kiefer (1959), we obtain the following table:

α	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
0.1	0.34730	0.60704	0.84116	1.06311	1.27748
0.05	0.46136	0.74752	1.00018	1.23730	1.46466
0.01	0.74346	1.07366	1.35861	1.62263	1.87215

Table 3: Upper quantiles c defined by $P(\int_0^1 \sum_{i=1}^d B_i^2(t)dt > c) = \alpha$.

The quantiles for arbitrary d can be computed using formulas developed in Section 4. of Kiefer (1959). There are two typos in formula (4.4) of Kiefer (1959). The non-obvious one is that the right-hand side of (4.4) must be multiplied by $2^{-(h-2)/4}$ to yield the entries in Table 3 of Kiefer (1959).

Similarly as for statistic Λ_n^*, it is at present not known if relations like (3.8) hold if the observations exhibit some form of weak dependence, as it is the case for GARCH models.

We also consider a bootstrap test like the one described above with Λ_n^* and λ_n^* replaced, respectively, by Δ_n^* and δ_n^*.

4 Simulation study

In this section we present the results of a simulations study intended to assess and compare the performance of the change point detection procedures introduced in the previous sections.

We consider the Data Generating Processes, henceforth DGP's, which satisfy

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha y_{t-1}^2. \tag{4.1}$$

Unless specified otherwise, the innovations ε_t are standard normal. We consider three DGP's which satisfy the null hypothesis and nine DGP's representing various change-point alternatives.

DGP 0a ARCH(1):

$$\omega = 0.1, \quad \beta = 0, \quad \alpha = 0.5.$$

DGP 0b GARCH(1,1):

$$\omega = 0.1, \quad \beta = 0.3, \quad \alpha = 0.3.$$

DGP 0c ARCH(1):

$$\omega = 0.1, \quad \beta = 0, \quad \alpha = 0.8.$$

DGP 1 GARCH (1,1) process with abrupt change–point in the middle of the sample (large changes in parameters, large change in unconditional variance):

$$\begin{aligned} \omega &= 0.1, \quad \beta = 0.3, \quad \alpha = 0.3 \text{ for } t = 1, \dots, [n/2] \quad (\sigma^2 = 0.25), \\ \omega &= 0.15, \quad \beta = 0.65, \quad \alpha = 0.25 \text{ for } t = [n/2]+1, \dots, n \quad (\sigma^2 = 1.5). \end{aligned}$$

DGP 2 GARCH(1,1) process with abrupt change–point in the middle of the sample (large changes in parameters, small change in unconditional variance):

$$\begin{aligned} \omega &= 0.1, \quad \beta = 0.3, \quad \alpha = 0.3 \text{ for } t = 1, \ldots, [n/2] \;\; (\sigma^2 = 0.25), \\ \omega &= 0.125, \quad \beta = 0.6, \quad \alpha = 0.1 \text{ for } t = [n/2]+1, \ldots, n \;\; (\sigma^2 = 0.4667). \end{aligned}$$

DGP 3 GARCH(1,1) process with change–point in the middle of the sample, such that the unconditional variance $\omega/(1-\alpha-\beta)$ remains unchanged ($\sigma^2 = 0.25$)

$$\begin{aligned} \omega &= 0.1, \quad \beta = 0.3, \quad \alpha = 0.3 \text{ for } t = 1, \ldots, [n/2], \\ \omega &= 0.15, \quad \beta = 0.25, \quad \alpha = 0.15 \text{ for } t = [n/2]+1, \ldots, n. \end{aligned}$$

DGP 4 Smooth transition GARCH(1,1) process,

$$\sigma_t^2 = \omega + \omega^* F(t,[n/2]) + (\beta + \beta^* F(t,[n/2]))\sigma_{t-1}^2 + (\alpha + \alpha^* F(t,[n/2]))y_{t-1}^2, \tag{4.2}$$

with

$$\begin{aligned} \omega &= 0.1, \quad \beta = 0.3, \quad \alpha = 0.3, \\ \omega^* &= 0.05, \quad \beta^* = 0.35, \quad \alpha^* = -0.05, \quad \gamma = 0.05, \end{aligned}$$

where $F(t,k) = (1+\exp(-\gamma(t-k)))^{-1}$, γ is a strictly positive parameter controlling the smoothness of the transition. If γ is large, this DGP reduces to **DGP 1**. We choose here a small value for γ, i.e. the transition between the two processes is smooth as illustrated in Figures 1, 2 and 3 which plot the transition for the parameters of **DGP 4**. **DGP 4** is of interest for economic variables, the transition of which is smooth, see e.g. Hagerud (1997), González-Rivera (1998) and Teräsvirta (1998).

DGP 5 GARCH(1,1): This DGP is similar to **DGP 1**, except that the change–point occurs in the first third of the sample:

$$\begin{aligned} \omega &= 0.1, \quad \beta = 0.3, \quad \alpha = 0.3 \text{ for } t = 1, \ldots, [n/3] \;\; (\sigma^2 = 0.25), \\ \omega &= 0.15, \quad \beta = 0.65, \quad \alpha = 0.25 \text{ for } t = [n/3]+1, \ldots, n \;\; (\sigma^2 = 1.5). \end{aligned}$$

DGP 6 GARCH(1,1): This DGP is similar to **DGP 1**, except that the change–point occurs in the first sixth of the sample:

$$\begin{aligned} \omega &= 0.1, \quad \beta = 0.3, \quad \alpha = 0.3 \text{ for } t = 1, \ldots, [n/6] \;\; (\sigma^2 = 0.25), \\ \omega &= 0.15, \quad \beta = 0.65, \quad \alpha = 0.25 \text{ for } t = [n/6]+1, \ldots, n \;\; (\sigma^2 = 1.5). \end{aligned}$$

DGP 7 GARCH(1,1) with constant parameters but with a change in the distribution of the innovations in the middle of the sample: the parameters are the same as in **DGP 0b**, but the distribution of the innovations ε_t changes as follows:

$$\begin{aligned} \varepsilon_t &\sim N(0,1) \text{ for } t = 1, \ldots, [n/2], \\ \varepsilon_t &\sim t(7) \text{ for } t = [n/2]+1, \ldots, n. \end{aligned}$$

DGP 8 GARCH(1,1) with constant parameters but with a change in the distribution of the innovations in the middle of the sample: the parameters are the same as in **DGP 0b** but the innovations ε_t for the second half of the sample follow a centralized and standardized $\chi^2(5)$ distribution:

$$\begin{aligned} \varepsilon_t &\sim N(0,1) \text{ for } t = 1, \ldots, [n/2], \\ \varepsilon_t &\sim \chi^2(5) \text{ for } t = [n/2]+1, \ldots, n. \end{aligned}$$

DGP 9 GARCH(1,1) with constant parameters but with a change in the distribution of the innovations in the middle of the sample: the parameters are the same as in **DGP 0b**, but the innovations ε_t for the second half of the sample follow a Laplace (two–sided exponential) distribution:

$$\begin{aligned} \varepsilon_t &\sim N(0,1) \text{ for } t = 1, \ldots, [n/2], \\ \varepsilon_t &\sim \text{Lap}(2^{-1/2}) \text{ for } t = [n/2]+1, \ldots, n. \end{aligned}$$

(The Laplace density, $\text{Lap}(\sigma)$, is $f(t) = (2\sigma)^{-1/2} \exp\{-|t|/\sigma\}$. If $\sigma = 2^{-1/2}$, its expected value is zero and the variance is one.)

The noise sequences in the DGP's are obtained by using the Box–Muller transformation of a sequence of uniform deviates which passes Marsaglia's (1996) DIEHARD randomness tests. The parameters are estimated by pseudo-maximum likelihood (PML) with analytic derivatives, see e.g. Bollerslev (1986). The consistency and asymptotic normality of the PML method for GARCH processes without finite fourth moments was recently verified by Berkes *et al.* (2003) where references to previous work are also given.

We consider two sample sizes, $n = 300$ and $n = 600$. Tables 4, 5 and 6 report simulation results. The first line reports the bootstrap tests, the second line the asymptotic tests: in the latter case, the average bootstrap P-value, $\overline{\hat{P}^*}$, is not reported.

Given that the computation of 5,000 replications of a bootstrap test with a preselection procedure for the number of bootstraps is time consuming, we restrict our simulations for $n = 600$ to statistics Λ_n^* and Δ_n^*, and consider fixed $B = 399$ for all replications.

To facilitate graphical comparison of the tests and to adjust the power of these tests to their size, we use size–power curves, i.e. the plots of the empirical distribution, henceforth EDF, of the P-values of the

DGP under the alternative against the EDF of the P–values of the DGP satisfying the null. We choose **DGP 0b** as the DGP satisfying the null hypothesis, as it is the closest, according to the Kullback–Leibler criterion, to the **DGP 1**, **DGP 2**, **DGP 3**, **DGP 4**, **DGP 5**, **DGP 6**, **DGP 7**, **DGP 8** and **DGP 9**, all of which follow the model of **DGP 0b** before a change–point.

For a given test, we define the EDF of the P-values $P_j, j = 1, \ldots, N$ as

$$\hat{F}(x_i) \equiv \frac{1}{N} \sum_{j=1}^{N} \mathbf{I}[P_j \leq x_i], \tag{4.3}$$

where $\{x_i\}$ is a fine grid of points in $[0, 1]$.

Figures 4, 5, 6, 7, 8, 9, 10, 11 and 12 display these curves for the 9 DGP's, together with the 45^o degree line.

Table 4: Empirical rejection probabilities for $n = 300$ for tests based on statistics $\hat{M}_n$ (K–S) and $\hat{B}_n$ (CVM) defined, respectively, by (2.10) and (2.13). For each DGP, the first row refers to a bootstrap test, the second to an asymptotic test. Each row is based on 5000 replications.

	CVM				K–S			
DGP	$\overline{\hat{P}^*}$	10 %	5 %	1 %	$\overline{\hat{P}^*}$	10 %	5 %	1 %
DGP 0a	0.5019	0.1035	0.0526	0.0159	0.4981	0.1041	0.0577	0.0175
DGP 0a	—	0.1039	0.0458	0.0098	—	0.1575	0.0878	0.0251
DGP 0b	0.5058	0.1001	0.0564	0.0159	0.5030	0.0999	0.0538	0.0141
DGP 0b	—	0.1059	0.0518	0.0072	—	0.1558	0.0854	0.0215
DGP 0c	0.4977	0.1020	0.0549	0.0144	0.4999	0.1032	0.0512	0.0065
DGP 0c	—	0.1042	0.0522	0.0067	—	0.1563	0.0879	0.0241
DGP 1	0.2254	0.4514	0.3236	0.1461	0.2301	0.4232	0.3083	0.1322
DGP 1	—	0.5218	0.3737	0.1316	—	0.5851	0.4365	0.2071
DGP 2	0.3141	0.3140	0.2074	0.0860	0.3336	0.2752	0.1824	0.0744
DGP 2	—	0.3164	0.1980	0.0562	—	0.3742	0.2452	0.0966
DGP 3	0.4957	0.1066	0.0580	0.0198	0.4953	0.1070	0.0582	0.0162
DGP 3	—	0.1040	0.0496	0.0110	—	0.1580	0.0834	0.0222
DGP 4	0.1753	0.5577	0.4211	0.2031	0.2126	0.4413	0.3099	0.1282
DGP 4	—	0.6152	0.4563	0.1979	—	0.6116	0.4393	0.2164
DGP 5	0.1783	0.5424	0.4163	0.1824	0.1962	0.4864	0.3472	0.1394
DGP 5	—	0.6041	0.4554	0.1826	—	0.6373	0.4894	0.2480
DGP 6	0.2385	0.4017	0.2694	0.1039	0.2964	0.2970	0.1794	0.0534
DGP 6	—	0.4534	0.2985	0.1018	—	0.4372	0.2910	0.1117
DGP 7	0.4054	0.2152	0.1335	0.0483	0.4096	0.2084	0.1272	0.0418
DGP 7	—	0.2070	0.1290	0.0403	—	0.2728	0.1779	0.0645
DGP 8	0.4046	0.1640	0.0836	0.0235	0.4194	0.1561	0.0855	0.0218
DGP 8	—	0.1550	0.0747	0.0167	—	0.2251	0.1245	0.0364
DGP 9	0.1298	0.6764	0.5642	0.3363	0.1397	0.6498	0.5323	0.3161
DGP 9	—	0.6683	0.5538	0.2984	—	0.7250	0.6127	0.3997

Table 5: Empirical rejection probabilities for $n = 300$ for tests based on statistics Λ_n^* (Λ) and Δ_n^* (Δ) defined, respectively, by (3.5) and (3.8). For each DGP, the first row refers to a bootstrap test, the second to an asymptotic test. Each row is based on 5000 replications.

	Λ				Δ			
DGP	$\overline{\hat{P}^*}$	10 %	5 %	1 %	$\overline{\hat{P}^*}$	10 %	5 %	1 %
DGP 0a	0.4900	0.1106	0.0668	0.0163	0.4935	0.1116	0.0644	0.0208
DGP 0a	—	0.0179	0.0055	0.0000	—	0.1159	0.0597	0.0128
DGP 0b	0.4909	0.1067	0.0585	0.0157	0.4821	0.1165	0.0627	0.0176
DGP 0b	—	0.0106	0.0024	0.0000	—	0.0680	0.0320	0.0038
DGP 0c	0.4979	0.1017	0.0504	0.0138	0.4924	0.1047	0.0579	0.0166
DGP 0c	—	0.0207	0.0043	0.0002	—	0.1012	0.0502	0.0079
DGP 1	0.0212	0.9458	0.8916	0.6980	0.0245	0.9378	0.8860	0.6908
DGP 1	—	0.9348	0.8215	0.3277	—	0.9582	0.9114	0.7076
DGP 2	0.1309	0.6510	0.5280	0.3204	0.1062	0.7070	0.5998	0.3838
DGP 2	—	0.3702	0.1906	0.0169	—	0.6584	0.5212	0.2532
DGP 3	0.4190	0.1856	0.1158	0.0434	0.3935	0.2102	0.1380	0.0582
DGP 3	—	0.0306	0.0096	0.0000	—	0.1320	0.0772	0.0214
DGP 4	0.0372	0.9009	0.8039	0.5186	0.0258	0.9407	0.8834	0.6318
DGP 4	—	0.8608	0.6607	0.1665	—	0.9583	0.9093	0.6875
DGP 5	0.0238	0.9421	0.8751	0.6206	0.0295	0.9287	0.8484	0.5721
DGP 5	—	0.9248	0.7948	0.3064	—	0.9419	0.8806	0.6620
DGP 6	0.0505	0.8613	0.7302	0.3400	0.1096	0.6685	0.4804	0.1587
DGP 6	—	0.7961	0.5892	0.1574	—	0.7043	0.5672	0.3066
DGP 7	0.3287	0.2729	0.1792	0.0707	0.3598	0.2263	0.1384	0.0412
DGP 7	—	0.0740	0.0367	0.0086	—	0.1487	0.0831	0.0230
DGP 8	0.3100	0.3005	0.1988	0.0824	0.3403	0.2439	0.1510	0.0456
DGP 8	—	0.0895	0.0420	0.0090	—	0.1664	0.0915	0.0237
DGP 9	0.2026	0.4694	0.3505	0.1724	0.2752	0.3383	0.2259	0.0888
DGP 9	—	0.1828	0.0909	0.0153	—	0.2436	0.1537	0.0506

Table 6: Empirical rejection probabilities for $n = 600$ for tests based on statistics Λ_n^* and Δ_n^* defined, respectively, by (3.5) and (3.8). For each DGP, the first row refers to a bootstrap test, the second to an asymptotic test. Each row is based on 5000 replications.

	Λ				Δ			
DGP	$\overline{\hat{P}^*}$	10 %	5 %	1 %	$\overline{\hat{P}^*}$	10 %	5 %	1 %
DGP 0a	0.4980	0.1058	0.0538	0.0122	0.4976	0.1080	0.0572	0.0142
DGP 0a	—	0.0257	0.0070	0.0002	—	0.1102	0.0576	0.0120
DGP 0b	0.4894	0.1086	0.0555	0.0143	0.4782	0.1111	0.0567	0.0122
DGP 0b	—	0.0209	0.0034	0.0000	—	0.0935	0.0437	0.0084
DGP 0c	0.4950	0.1048	0.0559	0.0122	0.4861	0.1103	0.0620	0.0153
DGP 0c	—	0.0285	0.0076	0.0005	—	0.1050	0.0577	0.0108
DGP 1	0.0014	0.9992	0.9972	0.9754	0.0018	0.9986	0.9944	0.9700
DGP 1	—	0.9994	0.9980	0.9316	—	0.9994	0.9976	0.9840
DGP 2	0.0329	0.9164	0.8642	0.7157	0.0271	0.9310	0.8873	0.7526
DGP 2	—	0.8307	0.6794	0.2665	—	0.9455	0.9018	0.7405
DGP 3	0.3411	0.2796	0.1865	0.0727	0.2991	0.3273	0.2296	0.0964
DGP 3	—	0.0792	0.0274	0.0021	—	0.2777	0.1834	0.0635
DGP 4	0.0028	0.9948	0.9885	0.9584	0.0025	0.9948	0.9911	0.9701
DGP 4	—	0.9988	0.9941	0.8899	—	0.9984	0.9968	0.9857
DGP 5	0.0015	0.9992	0.9980	0.9743	0.0023	0.9990	0.9973	0.9555
DGP 5	—	0.9992	0.9980	0.9291	—	0.9988	0.9980	0.9827
DGP 6	0.0120	0.9920	0.9706	0.6664	0.0322	0.9473	0.8313	0.3560
DGP 6	—	0.9918	0.9654	0.6982	—	0.9541	0.9064	0.7375
DGP 7	0.2954	0.3276	0.2190	0.0925	0.3409	0.2494	0.1561	0.0580
DGP 7	—	0.1185	0.0562	0.01345	—	0.2137	0.1338	0.0445
DGP 8	0.2693	0.3723	0.2664	0.1202	0.3143	0.2910	0.1862	0.0620
DGP 8	—	0.1499	0.0754	0.0157	—	0.2550	0.1575	0.0456
DGP 9	0.1668	0.5430	0.4154	0.2218	0.2534	0.3813	0.2621	0.1098
DGP 9	—	0.2655	0.1441	0.0229	—	0.3388	0.2308	0.0848

We now discuss the main conclusions that can be drawn from our simulations.

First of all, given that Chu (1995) used 300 observations, a direct comparison is possible. Except the K–S and Λ asymptotic tests which do not have the correct size, the remaining tests have better size and power than the ones considered by Chu (1995).

For the tests considered in this paper, the first conclusion is that the tests based on the generalized likelihood ratio, i.e. the tests Λ and Δ are generally more powerful that the tests based on the squared residuals. When considering a change in the distribution of the innovations in the middle of the sample, i.e. **DGP 7**, **DGP 8** and **DGP 9**, the discrepancy in power between these to classes of tests becomes smaller: while the Λ and Δ appear to be slightly more powerful for **DGP 7** and **DGP 8**, for **DGP 9** the statistics based on the empirical process are more powerful than the GLR based tests. This can be due to the fact that the Laplace distribution has much heavier tails than the normal distribution and so the residuals have very different distributions before and after the change. Another factor may be that the PML estimators are very biased as they are based on a Gaussian likelihood what again leads to very different residuals before and after the change. For **DGP 7**, **DGP 8** and **DGP 9**, bootstrap tests are more powerful than the asymptotic tests, the asymptotic Λ test appears to have very little power. The power of the GLR tests does not increase too much when the sample size increases from 300 to 600 observations.

The asymptotic test Λ has incorrect size and consequently its power is also smaller than that of the bootstrap test Λ. The bootstrap test Δ is comparable to the bootstrap test Λ and to the asymptotic test Δ. Consequently, we would recommend asymptotic test Δ if the calculation of the bootstrap test Δ is too time consuming.

The power of all tests is close to the size for **DGP 3**, for which the unconditional variance remains constant, see Figure 6. However the bootstrap test Δ has power of 23% even in this case when $n = 600$. Generally, and not surprisingly, the greater the change in unconditional variance, the easier it is to detect. This result is of practical interest as according to Mikosch and Stărică (2002b), long range dependence (LRD) in the processes $\{|y_t|\}$ and $\{y_t^2\}$ can be caused by a change in the unconditional variance of the process $\{y_t\}$, the intensity of the degree of LRD being an increasing function of the magnitude of the changes in the unconditional variances.

The power of the CVM and K–S tests increases when the transition is smooth, **DGP 4**, or occurs in the first third of the sample **DGP 5**. For these two DGP, the Λ test looses some power, while the test Δ is unaffected. We note that when the transition occurs near the beginning of the sample, i.e. for **DGP 6**, all statistics loose some power, the most affected being the weighted LR test Δ, and the K–S statistic. We also observe that for **DGP 6** asymptotic and bootstrap tests significantly differ, with the asymptotic test being more powerful. This reinforces our recommendation to use the asymptotic test Δ for large samples for which bootstrap based inference might be time consuming, even if this statistic has the lowest power for the extreme case of **DGP 9**.

5 Application to financial returns

We consider two groups of series of daily returns on financial assets. The data have been collected from Bloomberg. The first group contains series with a well known structural break: the British Pound/US Dollar FX rate, the S&P 500, FTSE 100 and NASDAQ indexes. Mikosch and Stărică (1999, 2002a) and Kokoszka and Leipus (2000), among others, have confirmed a change in regime in the first two series using different statistical methods. For the GBP–USD FX rate we consider samples of 300 observations starting from March 1979, while for the S&P 500 Composite index, FTSE 100 and NASDAQ, we consider 300 observations from June 1, 1987. For the FX rates, the considered period includes a structural break caused by the inception of the European Monetary System in April 1979, while for the S&P 500, FTSE 100 and NASDAQ indexes, a change in regime caused by the 19^{th} October 1987 "black Monday" stock market crash is likely.

The returns in the second group are observed from April 8, 1999 until August 2001. For these series, there is a priori no known structural break. This group contains returns on equities in the technology sector, i.e. IBM, Dell, Oracle, SAP, and Microsoft, and in the financial sector, i.e. BNP Paribas (BNP), Bank of America (BoA), Citygroup, HSBC, JP Morgan Chase (JPM) and Lloyds. For these series we consider two consecutive samples of 300 observations, i.e. BoA_1 denotes the first 300 observations for the BoA series, while BoA_2 denotes the subsequent 300 observations for that series.

We worked with the series of log returns $r_t = 100\ln(P_t/P_{t-1})$, where P_t denotes the asset price at time t. We estimated the following GARCH(1,1) model on the series $\{r_t\}$:

$$r_t = \mu + \psi\varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2), \tag{5.1}$$

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2. \tag{5.2}$$

where the parameter ψ takes into account the short range dependence in asset price returns caused by differences in closing times, see e.g. Campbell *et al.* (1997). For several series, the parameter ψ is not significant. Table 7 reports the estimated parameters for the series and the bootstrap P–values for the four tests. The test were performed by replacing the y_t considered in Sections 2 and 3 by the estimated ε_t in (5.1). We now discuss the main conclusions that can be drawn from Table 7.

The most general observation is that the likelihood ratio test are much more likely to reject than the tests based on residuals. This accords with the conclusions drawn in Section 4.

In accordance with earlier studies, there is overwhelming evidence for the presence of a change–point in the series of returns on the GBP-USD exchange rate. For the three indexes, the tests based on the likelihood ratio indicate a change–point, while the tests based on the empirical process of residuals largely

Table 7: GARCH(1,1) parameter estimates and P-values for the bootstrap tests.

Series	μ	ω	β	α	P_Λ	P_Δ	P_{CVM}	P_{K-S}
GBP–USD	0.0361	0.1203	0.5629	0.0401	0.0000	0.0000	0.0000	0.0000
S&P 500	0.0675	0.2572	0.6429	0.2771	0.0092	0.0065	0.4747	0.6263
FTSE 100	0.1348	0.2818	0.7191	0.1491	0.0000	0.0000	0.0744	0.1809
NASDAQ	0.0793	0.0392	0.7403	0.2089	0.0262	0.0967	0.5757	0.8788
IBM_1	1.0000e-13	0.8645	0.8334	0.0470	0.0100	0.0000	0.3535	0.4545
IBM_2	0.0342	0.5486	0.8071	0.1235	0.0000	0.0563	0.1083	0.2222
Dell_1	0.0859	0.9668	0.8458	0.0452	0.1809	0.1269	0.3838	0.6465
Dell_2	0.1589	0.8059	0.8999	0.0628	0.1880	0.6161	0.2161	0.2010
Oracle_1	-0.2672	0.7342	0.5948	0.2336	0.0000	0.0000	0.1979	0.4040
Oracle_2	0.5880	0.8362	0.8919	0.0637	0.0526	0.2626	0.7879	0.7475
SAP_1	1.0000e-13	0.4257	0.7714	0.1595	0.0000	0.0000	0.8889	0.6263
SAP_2	1.0000e-13	0.1505	0.7805	0.1458	0.8081	0.2362	0.4848	0.4343
Microsoft_1	1.0000e-13	0.8644	0.7256	0.2040	0.0000	0.0000	0.5656	0.5454
Microsoft_2	0.0331	0.7144	0.8079	0.1166	0.0926	0.1288	0.4848	0.2929
BNP_1	0.3039	0.6004	0.7296	0.0908	0.1527	0.1244	0.1704	0.5252
BNP_2	0.2805	0.4036	0.7481	0.1187	0.0100	0.0125	0.3737	0.5454
BoA_1	0.0471	0.3403	0.9217	0.0471	0.0000	0.0275	0.4040	0.2311
BoA_2	0.0379	0.4536	0.7812	0.1173	0.0044	0.0100	0.0100	0.0100
Citygroup_1	0.0911	0.5901	0.8826	0.0224	0.2222	0.2311	0.4949	0.4545
Citygroup_2	1.0000e-13	0.3373	0.8300	0.1082	0.2929	0.7677	0.0452	0.0626
HSBC_1	0.1118	0.2425	0.7941	0.1583	0.0000	0.3636	0.1207	0.2222
HSBC_2	0.0082	0.3314	0.8781	0.0440	0.0288	0.1960	0.5353	0.2828
Lloyds_1	1.0000e-13	0.3721	0.8743	0.0780	0.0911	0.1439	0.5556	0.4545
Lloyds_2	0.1631	0.4700	0.8236	0.0623	0.0100	0.0025	0.0100	0.0688
JPM_1	1.0000e-13	0.4459	0.8963	0.0385	0.2828	0.3333	0.3333	0.3333
JPM_2	1.0000e-13	0.4903	0.8394	0.1116	0.0100	0.0563	0.1909	0.2424

do not provide evidence for a change–point. Since the latter two tests have power less than 50% for the nominal size of 5%, see Table 4, we believe that a change–point is present in the three indexes.

Turning to the returns on shares in technology companies, we see that the P–values for the IBM stock are similar to those for the three indexes, so we similarly conclude that a change–point is present in both IBM series. We also observe that a change–point is likely to be present for the first subsample of the series

IBM, Oracle, SAP and Microsoft, while there is no evidence for a change point in the second subsample. The first subsample includes the period where the prices of speculative assets in the technological sector strongly increased. No change–point has been detected for Dell, which is not surprising as Dell shares are not considered by fund managers as speculative assets. The series SAP_1 exhibits a very interesting behavior: the P–values for the GLR tests are equal to zero, whereas the P–values for the CVM and K–S test are nearly equal to one. This might indicate that the strong rejections by the GLR tests is due not only to change–points but also possibly to other departures from the postulated GARCH(1,1) model.

For the banking sector series, a change–point in the volatility process is likely for the second subsample of the series BNP, BoA, Lloyds and JP Morgan. For Lloyds, the large discrepancy between the wavelet and the local Whittle estimates confirms the presence of a change–point in the volatility process. For the BNP case, after the initial fear that this bank would stay out of the consolidation game with Paribas, the stock price increased nicely until February 2002. The detected changes in the volatility of BoA might be the consequence of the uncertainty on the viability of the merging of Nationsbank and Bankamerica. The investment bank JP Morgan is the outcome of the merging of two risky banks, JP Morgan and Chase Manhattan. Investment banks are more affected than retail banks by global economic conditions, and are then more volatile. For HSBC, the Λ statistics suggest the presence of change–points which might be caused by the consequences of the Asian crisis in August 1998: HSBC had strong positions through Asia, but recovered from this crisis, and the HSBC stock price has been rising steadily until February 2001 and stabilised afterwards. For Citygroup, the CVM and K–S statistics detect a change–point for the second subsample. Since Citygroup is one of the most integrated and global financial service group that has ever existed, it is affected by the macroeconomic environment, and a change in regime is likely.

While the presence of long–range dependence (LRD) characteristics in absolute returns has been well documented, see Granger and Ding (1995, 1996), among others, some attention has recently focused on the observation that changes in the structure of a time series may make it behaving in many respects like a stationary LRD series, see e.g. Bos *et al.* (1999), Diebold and Inoue (2001), Granger and Hyung (1999) and Hsu (2000) who consider linear models, and Mikosch and Stărică (1999) and Giraitis *et al.* (2001) who include the GARCH model. In particular, tests for long memory will typically reject a short memory null in the presence of change points and vice versa.

While the discussion on whether it is more correct or useful to use stationary long memory models or change–point models is likely to continue for some time, some new light on this debate can be shed by using a wavelet based estimator developed by Abry and Veitch (1998) and Veitch and Abry (1999) which is able to estimate the self-similarity parameter of a series even if this series has nonstationarities like change–points in the mean or low order polynomial trends, see Abry *et al.* (2000, 2002) for background and the details. Thus, in a sense, this estimator, can isolate pure long memory from the spurious long memory caused by change–points and other nonstationarities. No formal tests exploiting this property have been developed yet, but we think it is worthwhile to apply this estimator to the series from Table 7 and compare it with the more traditional local Whittle estimator of Robinson (1995). Both estimators are based on the assumption that in a close positive neighborhood of the zero frequency, the spectrum of a long–memory process has the form:

$$f(\lambda) \sim c_f \lambda^{-\vartheta}, \quad \lambda \to 0_+, \tag{5.3}$$

where ϑ is the scaling parameter and c_f is a constant. The parameter ϑ is related to the fractional difference parameter by $\vartheta = 2d$.

The local Whittle estimator depends on a a bandwidth parameter m for which we used the data driven selection procedure of Henry and Robinson (1996), see also Henry (2001). We used a version of the wavelet estimator described in Veitch and Abry (1999) with Daubechies wavelets with four vanishing moments. We estimated the scaling parameter on the whole samples of 600 observations.

Table 8: Estimates of the scaling parameter ϑ. Asymptotic standard errors in parentheses.

	Local Whittle	Wavelet
GBP–USD	0.5916 (0.1387)	0.1260 (0.0919)
S&P 500	0.6048 (0.0619)	0.1451 (0.0919)
FTSE 100	0.6378 (0.1222)	0.1036 (0.0919)
NASDAQ	0.5875 (0.0816)	0.0927 (0.0919)
IBM	0.3628 (0.1104)	0.1770 (0.0919)
Dell	0.7470 (0.1387)	0.2050 (0.0919)
Oracle	0.6012 (0.1139)	0.1021 (0.0919)
SAP	0.5017 (0.1231)	0.3747 (0.0919)
Microsoft	0.6407 (0.1291)	0.1528 (0.0919)
BNP	0.3585 (0.1260)	0.1941 (0.0919)
BoA	0.2230 (0.0814)	0.1862 (0.0919)
Citygroup	0.6224 (0.1240)	0.2047 (0.0919)
HSBC	0.3122 (0.0962)	0.1980 (0.0919)
Lloyds	0.7470 (0.1387)	0.1366 (0.0919)
JPM	0.5174 (0.1104)	0.1332 (0.0919)

These results show a substantial discrepancy between the estimates obtained with the wavelet estimator and the local Whittle estimator, the scaling parameter ϑ being lower for the wavelet estimator. Furthermore, the 95-percent confidence intervals for the wavelet estimate ϑ often include the value zero. Thus, as indicated by the analysis in Table 7, the observed long range dependence in asset prices volatility might partially be the consequence of changes in the parameters of the volatility process. This accords with the models of Kirman and Teyssière (2002a, 2002b) in which long range dependence in asset prices volatility is linked to swings in opinion among heterogeneous market participants. In fact, in Kirman and Teyssière (2002b) such a discrepancy between the two estimators is observed in the absolute and squared returns generated by the model.

Acknowledgment. We thank Lajos Horváth for drawing our attention to the weighted statistic Δ_n^* defined in relation (3.8). We also thank Stelios Padelidakis and Stéphane Hochourian for providing useful hints for the analysis of the data in Section 5.

References

Abry, P., Flandrin, P., Taqqu, M. S. and Veitch, D. (2000). Wavelets for the analysis, estimation and synthesis of scaling data. In *Self-Similar Network Traffic and Performance Evaluation, New York* (eds K. Park and W. Willinger), pp. 39–88. Wiley (Interscience Division).

Abry, P., Flandrin, P., Taqqu, M. S. and Veitch, D. (2002). Self-similarity and long-range dependence through the wavelet lens. In *Theory and Applications of Long Range Dependence*, *Boston* (eds P. Doukhan, G. Oppenheim and M. S. Taqqu). Birkhäuser.

Abry, P. and Veitch, D. (1998). Wavelet analysis of long-range-dependent traffic. *IEEE Transactions on Information Theory*, **44,** number 1, 2–15.

Andreou, E. and Ghysels, E. (2001). Detecting multiple breaks in financial market volatility dynamics. Technical Report 2001s-65. CIRANO.

Bai, J. (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models. *The Annals of Statistics*, **22,** 2051–2061.

Berkes, I., Gombay, E., Horváth, L. and Kokoszka, P. S. (2002). Sequential change-point detection in GARCH(p, q) models. working paper available at `http://stat.usu.edu/~piotr/research.html`.

Berkes, I. and Horváth, L. (2002). Limit results for the empirical process of squared residuals in GARCH models. preprint.

Berkes, I., Horváth, L. and Kokoszka, P. S. (2002). Asymptotics for GARCH squared residual correlations. *Econometric Theory*; forthcoming, preprint available at `http://math.usu.edu/~piotr/research.html`.

Berkes, I., Horváth, L. and Kokoszka, P. S. (2003). GARCH processes: structure and estimation. *Bernoulli*; forthcoming, preprint available at `http://math.usu.edu/~piotr/research.html`.

Besseville, M. and Nikifirov, I. V. (1993). *Detection of Abrupt Changes: Theory and Applications.* Prentice Hall, Upper Saddle River.

Bickel, P. J. and Wichura, M. J. (1971). Convergence of multiparameter stochastic processes and some applications. *Ann. Math. Statist.*, **42,** 1656–1670.

Blum, J. R., Kiefer, J. and Rossenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, **32,** 485–498.

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

Bos, C., Franses, P. H. and Ooms, M. (1999). Long-memory and level shifts: re-analyzing inflation rates. *Empirical Economics*, **24,** 427–449.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets.* Princeton University Press, New Jersey.

Chu, C.-S. J. (1995). Detecting parameter shift in GARCH models. *Econometric Reviews*, **14,** 241–266.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* Wiley, New York.

Davidson, R. and MacKinnon, J. G. (1998). Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School*, **66,** 1–26.

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

Diebold, F. and Inoue, A. (2001). Long memory and structural change. *Journal of Econometrics*, **105,** 131–159.

Giraitis, L., Kokoszka, P. S. and Leipus, R. (2001). Testing for long memory in the presence of a general trend. *Journal of Applied Probability*, **38,** 1033–1054.

González-Rivera, G. (1998). Smooth transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, **3,** 61–78.

Granger, C. W. and Hyung, N. (1999). Occasional structural breaks and long memory. Discussion paper 99-14. University of California, San Diego. Available at `http://www.econ.ucsd.edu/papers/`.

Granger, C. W. J. and Ding, Z. (1995). Some properties of absolute returns, an alternative measure of risk. *Annales d'Économie et de Statistique*, **40,** 67–91.

Granger, C. W. J. and Ding, Z. (1996). Varieties of long-memory models. *Journal of Econometrics*, **73,** 61–77.

Hagerud, G. E. (1997). A new non-linear GARCH model. Working Paper Series in Economics and Finance 162. Stockholm School of Economics.

Henry, M. (2001). Robust automatic bandwidth for long-memory. *Journal of Time Series Analysis*, **22,** 293–316.

Henry, M. and Robinson, P. M. (1996). Bandwidth choice in gaussian semiparametric estimation of long range dependence. In *Athens Conference on Applied Probability and Time Series Analysis, In Memory of E. J. Hannan, New York* (eds P. M. Robinson and M. Rosenblatt), pp. 220–232. Springer Verlag, LNS 115.

Horváth, L., Kokoszka, P. S. and Teyssière, G. (2001). Empirical process of the squared residuals of an ARCH sequence. *The Annals of Statistics*, **29,** 445–469.

Horváth, L., Kokoszka, P. S. and Teyssière, G. (2002). Bootstrap specification tests for ARCH based on the empirical process of squared residuals. Tentatively accepted by Journal of Statistical Computation and Simulation. Preprint available at `http://www.math.usu.edu/~piotr/research.html`.

Hsu, C-C (2000). Long memory or structural change: testing method and empirical examination. Preliminary version.

Hull, J. C. (2000). *Options, Futures & other Derivatives.* Prentice Hall, Upper Saddle River, NJ.

Kiefer, J. (1959). K-sample analogues of the Kolmogorov-Smirnov and Cramér-v.Mises tests. *Ann. Math. Statist.*, **30,** 420–447.

Kirman, A. and Teyssière, G. (2002a). Microeconomic models for long-memory in the volatility of financial time series. *Studies in Nonlinear Dynamics and Econometrics*, **5,** 281–302.

Kirman, A. and Teyssière, G. (2002b). Testing for bubbles and change–points. Preprint.

Kokoszka, P. S. and Leipus, R. (1998). Change–point in the mean of dependent observations. *Statistics and Probability letters*, **40,** 385–393.

Kokoszka, P. S. and Leipus, R. (1999). Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal*, **39,** 231–247.

Kokoszka, P. S. and Leipus, R. (2000). Change-point estimation in ARCH models. *Bernoulli*, **6,** 513–539.

Kokoszka, P. S. and Leipus, R. (2002). Detection and estimation of changes in regime. In *Theory and Applications of Long-range Dependence*, *Boston* (eds P. Doukhan, G. Oppenheim and M. S. Taqqu), pp. 325–337. Birkhäuser.

Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in time series. *Journal of Time Series Analysis*, **20,** 36–60.

Lundbergh, S. and Teräsvirta, T. (2002). Evaluating GARCH models. *Journal of Econometrics*, **110,** 417–435.

Mantalos, P. (2000). A graphical investigation of the size and power of the Granger causality test in integrated-cointegrated VAR systems. *Studies in nonlinear dynamics and econometrics*, **4,** 17–33.

Marsaglia, G. (1996). DIEHARD: A battery of tests of randomness. Available at `http://stat.fsu.edu/pub/diehard`.

McCullough, B. D. and Vinod, H. D. (1999). The numerical reliability of econometric sofware. *Journal of Economic Literature*, **37,** 633–665.

Mikosch, T. and Stărică, C. (1999). Change of structure in financial time series, long range dependence and the GARCH model. Technical Report. University of Groningen. preprint available at `http://www.math.ku.dk/~mikosch/preprint.html`.

Mikosch, T. and Stărică, C. (2002a). Long-range dependence effects and ARCH modeling. In *Theory and Applications of Long-range Dependence*, *Boston* (eds P. Doukhan, G. Oppenheim and M. S. Taqqu), pp. 439–459. Birkhäuser.

Mikosch, T. and Stărică, C. (2002b). Non-stationarities in financial time series: the long range dependence and the IGARCH effects. Technical Report. University of Copenhagen.

Picard, D. (1985). Testing and estimating change–point in time series. *Adv. in Appl. Probab.*, **17,** 841–867.

Rice, J. A. (1995). *Mathematical Statistics and Data Analysis*, Second edn. Duxbury Press.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, **23,** 1630–1661.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

Stock, J. H. and Watson, M. W. (2002). Has the business cycle changed and why ? Technical Report. Harvard University.

Teräsvirta, T. (1998). Modelling economic relationships with smooth transition regressions. In *Handbook of Applied Economic Statistics* (eds A. Ullah and D. E. A. Giles), pp. 507–552. Marcel Dekker.

Veitch, D. and Abry, P. (1999). A wavelet based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory special issue on "Multiscale Statistical Signal Analysis and its Applications"*, **45,** number 3, 878–897.

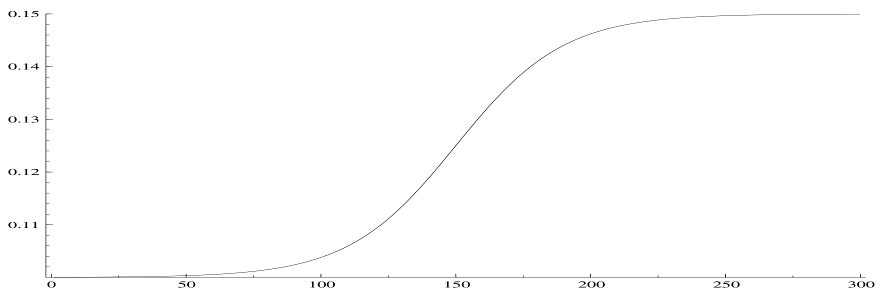

Figure 1: Parameter $\omega + \omega^{*}F(\cdot)$ of the smooth transition GARCH, DGP4

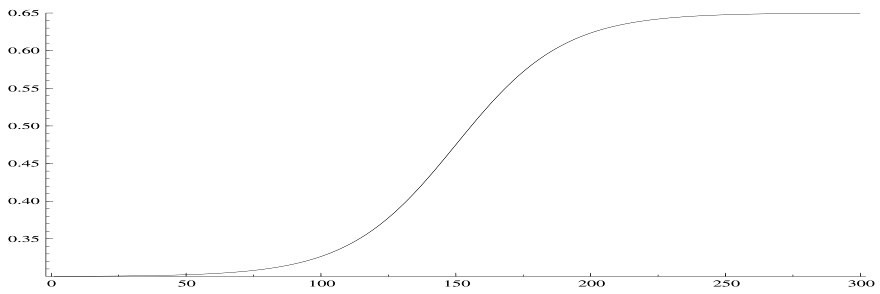

Figure 2: Parameter $\beta + \beta^{*}F(\cdot)$ of the smooth transition GARCH, DGP4

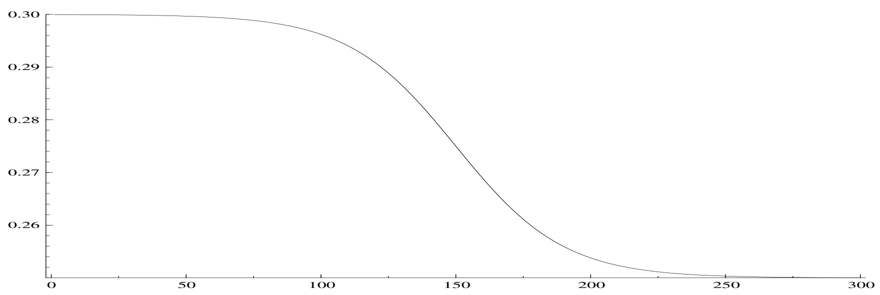

Figure 3: Parameter $\alpha + \alpha^{*}F(\cdot)$ of the smooth transition GARCH, DGP4

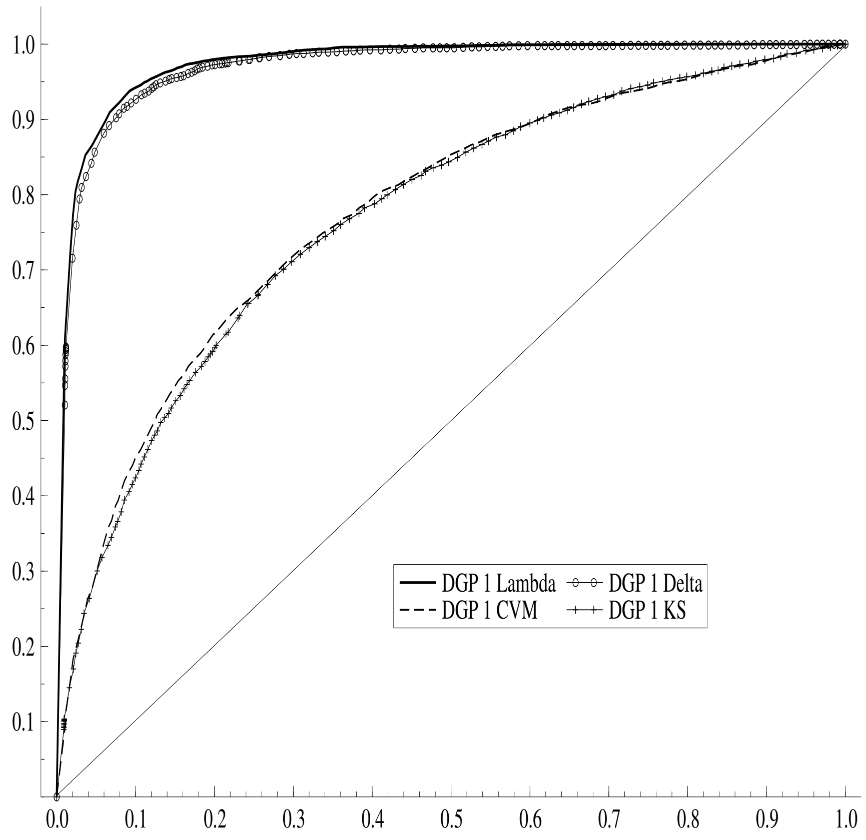

Figure 4: Size–Power Curves. DGP 1. 300 Observations

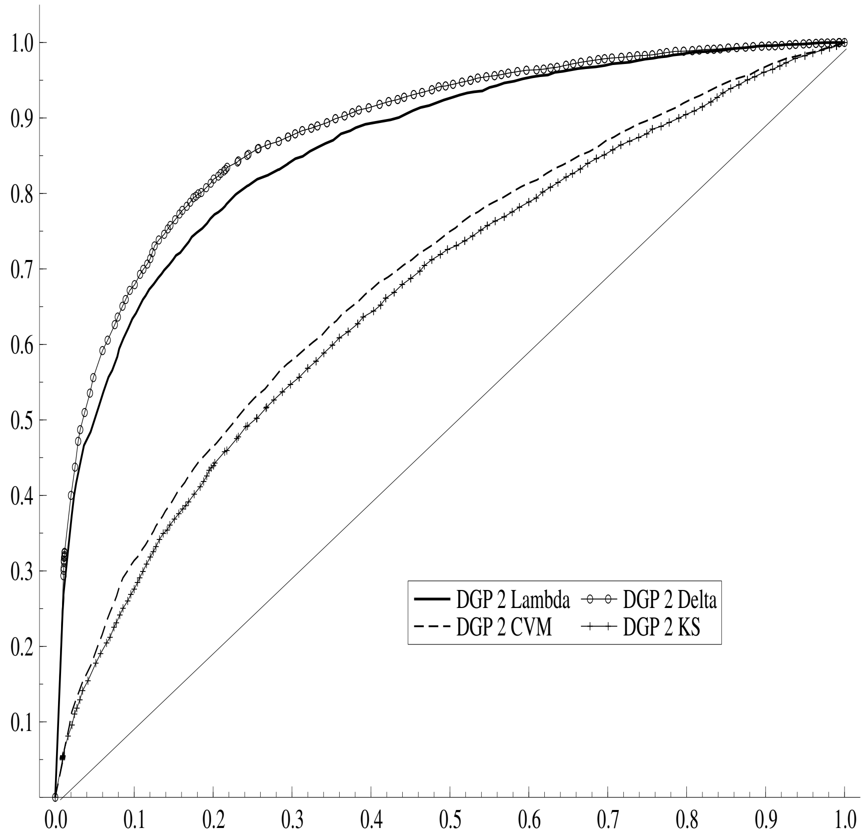

Figure 5: Size–Power Curves. DGP 2. 300 Observations

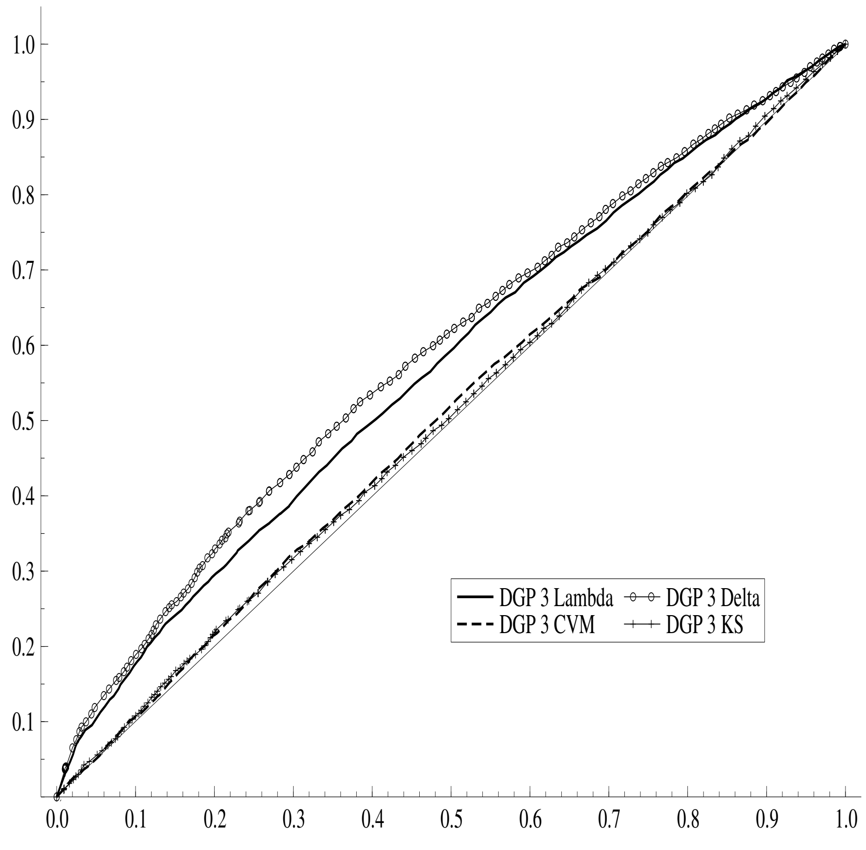

Figure 6: Size–Power Curves. DGP 3. 300 Observations

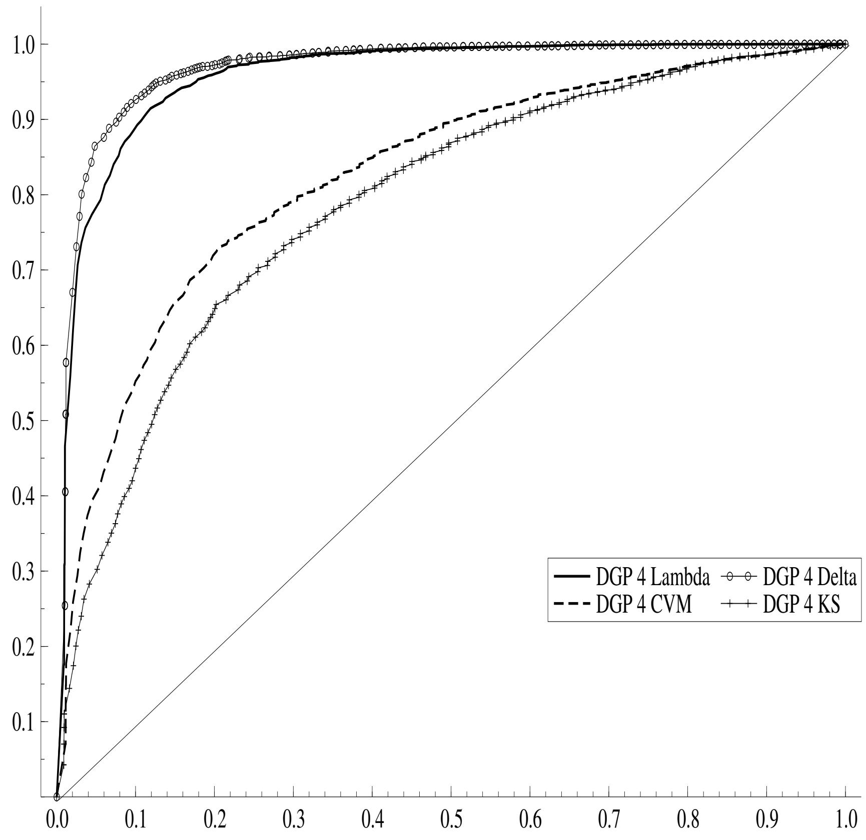

Figure 7: Size–Power Curves. DGP 4. 300 Observations

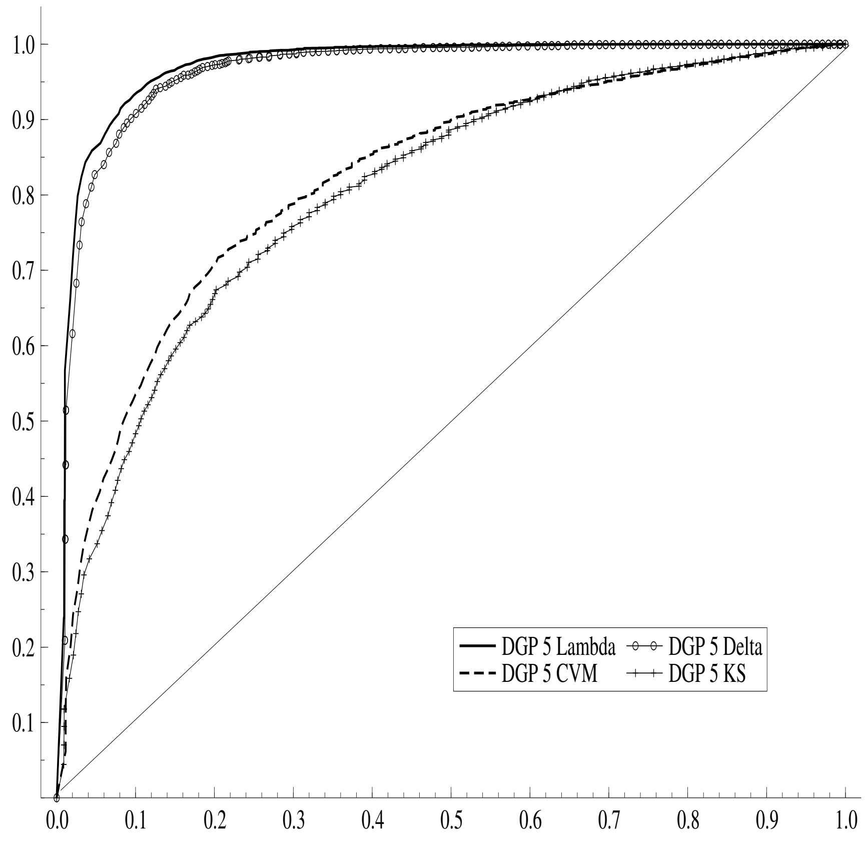

Figure 8: Size–Power Curves. DGP 5. 300 Observations

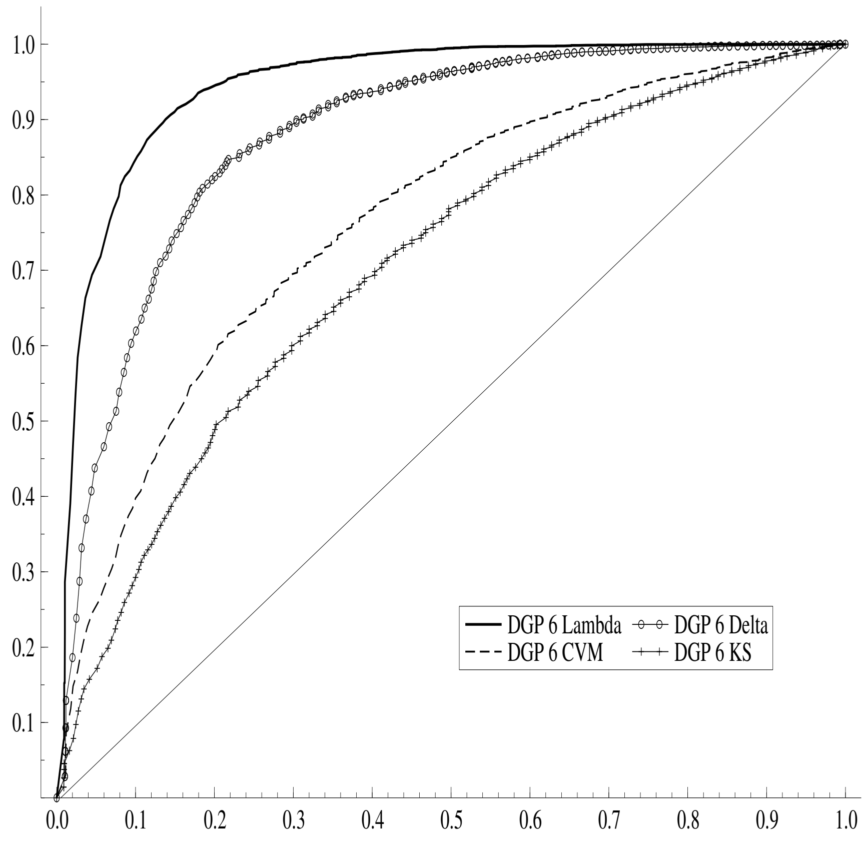

Figure 9: Size–Power Curves. DGP 6. 300 Observations

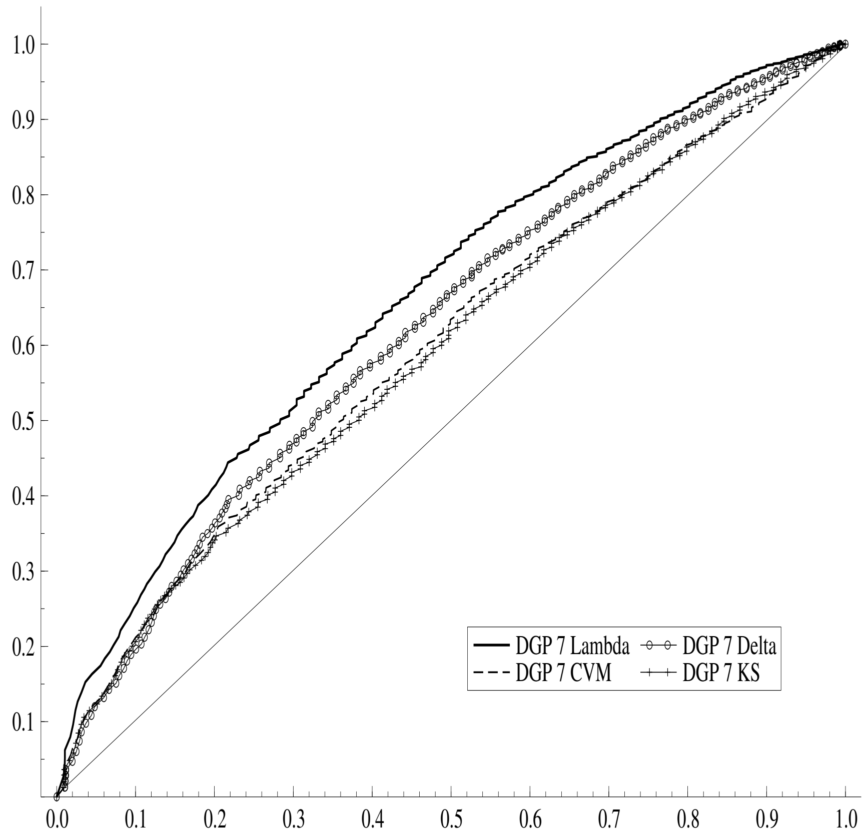

Figure 10: Size–Power Curves. DGP 7. 300 Observations

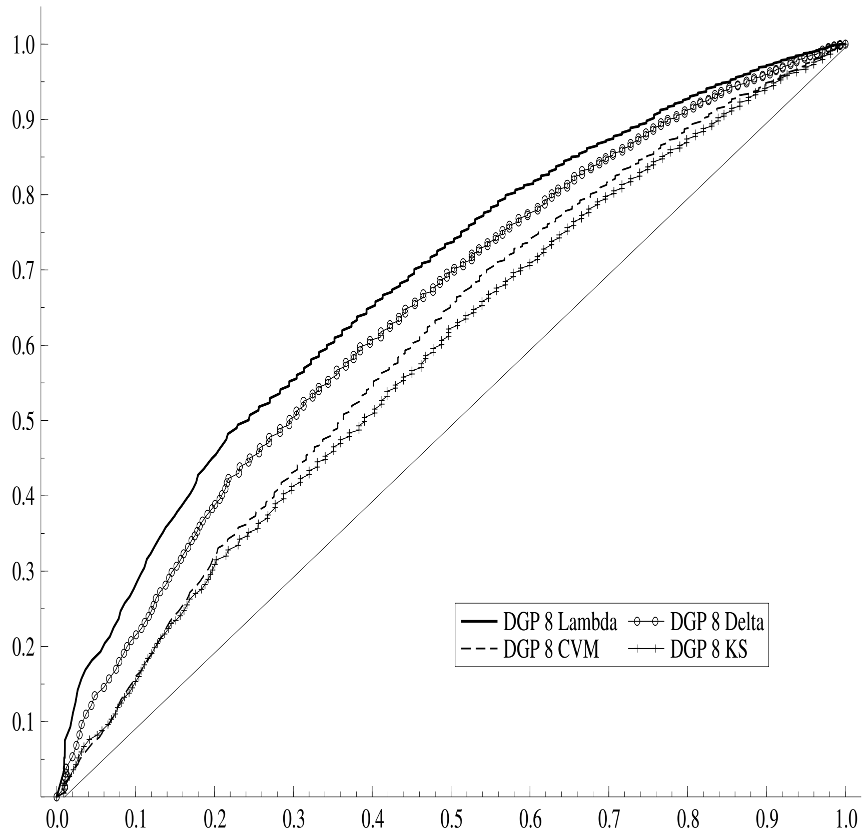

Figure 11: Size–Power Curves. DGP 8. 300 Observations

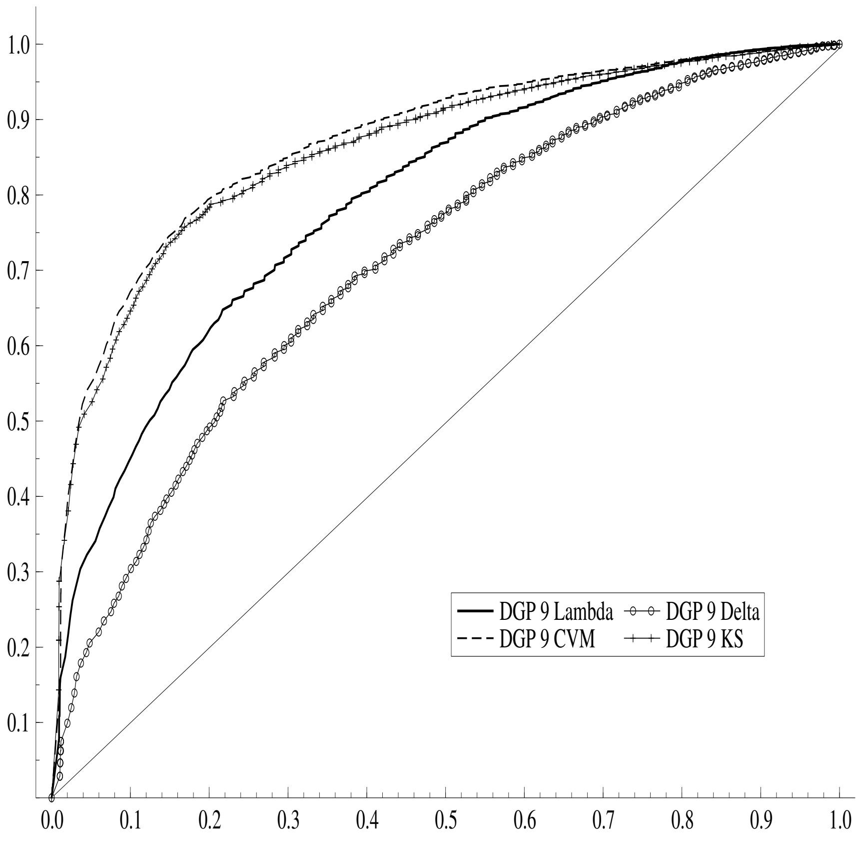

Figure 12: Size–Power Curves. DGP 9. 300 Observations