Testing for change in mean of heteroskedastic time series

Mohamed BOUTAHAR *

University of Méditerranée, Marseille, France

February 26, 2011

Abstract.

In this paper we consider a Lagrange Multiplier-type test (LM) to detect change in the mean of time series with heteroskedasticity of unknown form. We derive the limiting distribution under the null, and prove the consistency of the test against the alternative of either an abrupt or smooth changes in the mean. We perform also some Monte Carlo simulations to analyze the size distortion and the power of the proposed test. We conclude that for moderate sample size, the test has a good performance. We finally carry out an empirical application using the daily closing level of the S&P 500 stock index, in order to illustrate the usefulness of the proposed test.

AMS classifications codes: 62G10, 62G20, 60F17, 62M10

Keywords. Brownian bridge, changes in mean, functional central limit theorem, heteroskedasticity, time series

1 Introduction

In the statistic literature there is a vast amount of works on detecting changes in mean of a given time series. In a more general context of linear regression model, Chow (1960) considered tests for structural change for a known single break date. The researches headed for the modelling where this break date is treated as an unknown variable. Quandt (1960) extends the Chow test and proposes taking the largest Chow statistic over all possible break

---

*Departement of Mathematics, case 901. Luminy Faculty of Sciences. 163 Av. de Luminy 13288 Marseille Cedex 9. GREQAM, 2 rue de la charité 13002. e-mail: boutahar@univmed.fr.

dates. In the same context, the most important contributions are those of An-
drews (1993) and Andrews and Ploberger (1994). Sen and Srivastava (1975a,
1975b), Hawkins (1977), Worsley (1979), Srivastava and Worsley (1986) and
James et al. (1987) consider tests for mean shifts of normal sequence of vari-
ables. The multiple structural changes case receives an increasing attention.
For instance, Yao (1988), Yin (1988) and Yao and Au (1989) study the esti-
mation of the number of mean shifts of variables sequence using the Bayesian
information criterion. Liu et al. (1997) consider multiple changes in a linear
model estimated by least squares and estimate the number of changes using
a modified Schwarz' criterion. Bai and Perron (1998) consider the estimation
of multiple structural shifts in a linear model estimated by least squares; Qu
and Perron (2007) extend Bai and Perron's (1998) results to a multivariate
regression. In all these papers, a Wald, Lagrange Multiplier (LM) or/and
Likelihood-ratio (LR)-Like tests have been considered. Recall that the Wald
test is based on the unrestricted model, the LR test needs the restricted and
unrestricted model, while the LM test is based exclusively on the restricted
model.

Concerning only the change in mean, all authors cited above assume that
under the alternative hypothesis, the mean μ_t is a step function i.e. the
observations (y_t), $1 \leq t \leq n$, satisfy

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t, \\
\mu_t &= \mu_{(j)} \text{ if } t = n_{j-1} + 1, ..., n_j, n_j = [\lambda_j n], 0 < \lambda_1 < ... < \lambda_m < 1,
\end{aligned}
$$

where (ε_t) is such that $\mathrm{E}(\varepsilon_t) = 0$ and $[x]$ is the integer part of x. If the mean
μ_t is time varying with unknown form, then the Wald and LR tests can't be
applied. Only the LM test can be used since no specification of alternative
hypothesis is needed to build a statistic. Recently, Gombay (2008) used an
LM-type test for detecting change in the autoregressive model. However, he
assumed that the errors ε_t are homoskedastic i.e $\mathrm{var}(\varepsilon_t) = \sigma^2$ for all t. In
this paper we consider the heteroskedastic time series,

$$
y_t = \mu_t + \sigma_t \varepsilon_t, \tag{1}
$$

where the errors are Gaussian white noise $\varepsilon_t \sim N(0, 1)$, (σ_t) is a deterministic
sequence with unknown form. The null and the alternative hypotheses are
as follows:

$$
\left\{
\begin{array}{l}
H_0 : \mu_t = \mu \text{ for all } t \geq 1 \\
\text{against} \\
H_1 : \text{There exist } t \neq s \text{ such that } \mu_t \neq \mu_s
\end{array}
\right. , \tag{2}
$$

under the alternative hypothesis the mean μ_t can be time varying with un-
known form. The model (1) is useful in many areas. In financial modelling,

much research has been devoted to the study of long-run behavior of returns of speculative asset. A common finding in much of the empirical literature is that the returns are not serially correlated which is in agreement with the efficient market hypotheses, see Ding et al. (1993). However, the absolute returns which, is a proxy of the instantaneous standard deviation, has significant positive autocorrelations with a possible breaks in the mean and in the unconditional variance. For instance, Starica and Granger (2005) show that an appropriate model to describe the dynamic of the logarithm of the absolute returns of the S&P 500 index is given by (1) where μ_t and σ_t are step functions i.e.

$$
\begin{aligned}
\mu_t &= \mu_{(j)} \text{ if } t = n_{j-1} + 1, ..., n_j, n_j = [\lambda_j n], 0 < \lambda_1 < ... < \lambda_{m_1} < 1, \quad (3) \\
\sigma_t &= \sigma_{(j)} \text{ if } t = t_{j-1} + 1, ..., t_j, t_j = [\tau_j n], 0 < \tau_1 < ... < \tau_{m_2} < 1, \quad (4)
\end{aligned}
$$

for some integers m_1 and m_2. They also show that model (1), (3) and (4) gives forecasts superior to those based on a stationary GARCH(1,1) model.

One can also consider a more general model than (1), (3) and (4), where breaks can be abrupt and/or smooth. A model with $(m+1)$ regimes for the unconditional standard deviation can be defined by

$$
\sigma_t = \sum_{j=1}^m \sigma_{(j)} \left(1 - F_j \left(\frac{t/n - \tau_{2j-1}}{s_j} \right) \right) + \sigma_{(j+1)} F_j \left(\frac{t/n - \tau_{2j-1}}{s_j} \right), \quad (5)
$$

where $\tau_0 = 0 < \tau_1 < ... < \tau_{2m} < \tau_{2m+1} = 1$, F_j is the transition function from regime j to regime $(j+1)$, assumed to be continuous from $\mathbb{R}$ onto $[0, 1]$. The scale $s_j > 0$ indicates how rapidly the transition from regime j to regime $(j+1)$, a small s_j yields an abrupt change,

As in Gombay (2008) we use an LM-type test for detecting change in mean. The test statistic is based on the normalized score vector evaluated under the null $H_0' : \mu_t = \mu$ and $\sigma_t = \sigma$ for all $t \geq 1$. If $y_t = \mu + \sigma \varepsilon_t$ then the log-likelihood of the sample is given by

$$
L(n, \mu, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \mu)^2.
$$

Hence, the score vector is

$$
S_n(\mu, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} \sum_{t=1}^n (y_t - \mu) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^n (y_t - \mu)^2 \end{pmatrix}
$$

3

and the information matrix is $I_n(\mu, \sigma^2) = nI/\sigma^2$, I is the identity matrix. Therefore a test statistic for testing change in mean is based on the first component of the vector $I_n^{-1/2}(\widehat{\mu}, \widehat{\sigma}^2) S_{[n\tau]}(\widehat{\mu}, \widehat{\sigma}^2)$, where $\widehat{\mu} = \sum_{t=1}^n y_t/n$ and $\widehat{\sigma}^2 = \sum_{t=1}^n (y_t - \widehat{\mu})^2/n$ are the maximum likelihood estimators of μ and σ^2, given by

$$B_n(\tau) = \frac{1}{\sqrt{n}\widehat{\sigma}} \sum_{t=1}^{[n\tau]} (y_t - \widehat{\mu}).$$

The test statistic we consider is

$$\mathcal{B}_n = \sup_{\tau \in [0,1]} |B_n(\tau)|.$$

2 Limiting distribution of $\mathcal{B}_n$ under the null

Theorem 1. *Assume that (y_t) satisfies the model (1) with standard Gaussian white noise errors (ε_t) and a bounded deterministic sequence (σ_t) satisfying*

$$\frac{1}{n} \sum_{t=1}^n \sigma_t^2 \to \overline{\sigma}_2^2 \ as \ n \to \infty, \tag{6}$$

Then, under H_0 we have

$$\mathcal{B}_n \xrightarrow{\quad} \mathcal{B}_\infty = \sup_{\tau \in [0,1]} |B(\tau)| \tag{7}$$

$\xrightarrow{\mathcal{L}}$ *denotes the convergence in distribution and $B(\tau)$ is a Brownian Bridge.*

Remark. The condition (6) is a classical ergodic assumption and holds in many situations. For example if (σ_t) is given by (4), then (6) is satisfied with $\overline{\sigma}_2^2 = \sum_{j=1}^{m_2+1} (\tau_j - \tau_{j-1}) \sigma_{(j)}^2$, $\tau_0 = 0$, $\tau_{m_2+1} = 1$, and if (σ_t) is given by (5) then

$$\overline{\sigma}_2^2 = \int_0^1 \left\{ \sum_{j=1}^m \sigma_{(j)} \left(1 - F_j \left(\frac{x - \tau_{2j-1}}{s_j} \right) \right) + \sigma_{(j+1)} F_j \left(\frac{x - \tau_{2j-1}}{s_j} \right) \right\}^2 dx.$$

The proof of Theorem 1 is given in the Appendix.

3 Consistency of $\mathcal{B}_n$

3.1 Consistency of $\mathcal{B}_n$ against abrupt changes

Without loss of generality we assume that under the alternative hypothesis there is a single break date, i.e. (y_t) is given by (1) where

4

$$\mu_t = \begin{cases} \mu_{(1)} & \text{if } 1 \le t \le [n\tau_1] \\ \mu_{(2)} & \text{if } [n\tau_1] + 1 \le t \le n \end{cases} \qquad \text{for some } \tau_1 \in (0, 1). \tag{8}$$

Theorem 2. *Assume that (y_t) satisfies the model (1) with standard Gaussian white noise errors (ε_t) and a bounded deterministic sequence (σ_t) satisfying (6). If under H_1 the mean μ_t follows the dynamic (8) then*

$$\mathcal{B}_n \xrightarrow{\quad P \quad} + \infty \tag{9}$$

where $\xrightarrow{\quad P \quad}$ denotes the convergence in probability.

The proof of Theorem 2 is given in the Appendix.

Remark 1. The result of Theorem 2 remains valid if under the alternative hypothesis there are multiples breaks in the mean.

3.2 Consistency of $\mathcal{B}_n$ against smooth changes

In economics and finance, multiple regimes modelling becomes more and more important in order to take into account phenomena characterized, for instance, by recession or expansion periods, or high or low volatility periods. Consequently, it's more realistic to assume that the break in the mean doesn't happen suddenly but the transition from one regime to another is continuous with slowly variation. A well known dynamic is the smooth transition autoregressive (STAR) specification, see Terasvirta [22], in which the mean μ_t is a time varying with respect to the following

$$\mu_t = \mu_{(1)} + (\mu_{(2)} - \mu_{(1)})F(t/n, \tau_1, \gamma), 1 \le t \le n, \tag{10}$$

where $F(x, \tau_1, \gamma)$ is a the smooth transition function assumed to be continuous from $[0, 1]$ onto $[0, 1]$. The parameters $\mu_{(1)}$ and $\mu_{(2)}$ are the values of the mean in the two extreme regimes, that is when $F \to 0$ and $F \to 1$. The slope parameter γ indicates how rapidly the transition between two extreme regimes is. The parameter c is the location parameter.

Two choices for the function F are frequently evoked, the logistic function given by

$$F_L(x, \tau_1, \gamma) = [1 + \exp(-\gamma(x - \tau_1))]^{-1} \tag{11}$$

and the exponential one

$$F_e(x, \tau_1, \gamma) = 1 - \exp(-\gamma(x - \tau_1)^2). \tag{12}$$

For example, for the logistic function with $\gamma > 0$, the extreme regimes are obtained as follows

5

- if $x \to 0$ and γ large we have $F \to 0$ and thus $\mu_t = \mu_{(1)}$,
- if $x \to 1$ and γ large we have $F \to 1$ and thus $\mu_t = \mu_{(2)}$.

Theorem 3. *Assume that (y_t) satisfies (1) with standard Gaussian white noise errors (ε_t) and a bounded deterministic sequence (σ_t) satisfying (6). If under H_1 the mean μ_t follows the dynamic (10) then*

$$\mathcal{B}_n \xrightarrow{\ \ P\ \ } +\infty. \tag{13}$$

The proof of Theorem 3 is given in the Appendix.

4   Finite sample performance

All sequences are driven by a Gaussian white noise $\varepsilon_t \sim N(0,1)$. Simulations were performed using the software R [17]. We carry out an experiment of 1000 samples for nine series and we use four different sample sizes, $n = 30$, $n = 100$, $n = 500$ and $n = 1000$.

In the model (1) we consider three dynamics for the mean μ_t and the variance σ_t^2 :
- **Dynamics of the mean:**

$$\mu_t = 1 \text{ for all } t \geq 0, \tag{14}$$

$$\mu_t = \begin{cases} \mu_{(1)} & \text{if } 1 \leq t \leq [n\tau_1] \\ \mu_{(2)} & \text{if } [n\tau_1] + 1 \leq t \leq n \end{cases}, \tag{15}$$

$$\mu_t = \mu_{(1)} + (\mu_{(2)} - \mu_{(1)})F(t/n, \tau_1, \gamma), \tag{16}$$

we choose $\tau_1 = 0.5$ (one break in the middle of the sample), $\mu_{(1)} = 1, \mu_{(2)} = 2$, F is the logistic function given by (11) and $\gamma = 20$.
- **Dynamics of the variance**

$$\sigma_t = 1 \text{ for all } t \geq 0, \tag{17}$$

$$\sigma_t = \begin{cases} \sigma_{(1)} & \text{if } 1 \leq t \leq [n\tau_2] \\ \sigma_{(2)} & \text{if } [n\tau_2] + 1 \leq t \leq n \end{cases} \tag{18}$$

$$\sigma_t = \sigma_{(1)} + (\sigma_{(2)} - \sigma_{(1)})F(t/n, \tau_2, \gamma), \tag{19}$$

6

we choose $\tau_2 = 2/3$, $\sigma_{(1)} = 0.5$, $\sigma_{(2)} = 1.5$, F is the logistic function given by (11) and $\gamma = 20$.

To study the size of the test we simulate the following three series:

Series 1: μ_t is given by (14) and σ_t is given by (17), no break in the mean and in the variance.

Series 2: μ_t is given by (14) and σ_t is given by (18), no break in the mean and one abrupt break in the variance.

Series 3: μ_t is given by (14) and σ_t is given by (19), no break in the mean and one smooth break in the variance.

To study the power of the test we simulate the following six series:

• One abrupt change on the mean:

Series 4: μ_t is given by (15) and σ_t is given by (17), one abrupt break in the mean and no break in the variance.

Series 5: μ_t is given by (15) and σ_t is given by (18), one abrupt break in the mean and one abrupt break in the variance.

Series 6: μ_t is given by (15) and σ_t is given by (19), one abrupt break in the mean and one smooth break in the variance.

• A smooth change in the mean:

Series 7: μ_t is given by (16) and σ_t is given by (17), one smooth break in the mean and no break in the variance.

Series 8: μ_t is given by (16) and σ_t is given by (18), one smooth break in the mean and one abrupt break in the variance.

Series 9: μ_t is given by (16) and σ_t is given by (19), one smooth break in the mean and one smooth break in the variance.

Table 1. Empirical test sizes (in %)

	n α	$n = 30$	$n = 100$	$n = 500$	$n = 1000$
	1%	0.2	0.4	0.7	0.5
Series 1	5%	2.9	3.3	3.8	4.1
	10%	5.1	7.9	8.2	8.4
	1%	0.3	0.9	1.3	1.3
Series 2	5%	3.4	5.1	6.2	6.3
	10%	7.1	10.6	11.7	12.4
	1%	0.5	0.9	1.1	1.1
Series 3	5%	4.3	4.9	6.4	6.3
	10%	7.9	10.1	12.7	12.4

Note: Table 1 contains rejection frequencies of the null hypothesis of no change in the mean. Rejection frequencies are based on 1000 replications generated from the *Series* 1-3 where the nominal significance levels are 1% , 5% and 10%, the sample sizes are $n = 30, n = 100, n = 500$ and $n = 1000$.

Table 1 indicates that the test is a somewhat conservative (the empirical size is lesser than the nominal one) when the time series is homoskedastic (*Series* 1) and overrejects the null (the empirical size is greater than the nominal one) if the time series is heteroskedastic (*Series* 2 and 3).

Table 2. Empirical test powers (in %)

	n	$n = 30$	$n = 100$	$n = 500$	$n = 1000$
	α				
	1%	18.3	95.9	100	100
Series 4	5%	47.3	98.8	100	100
	10%	61.9	99.4	100	100
	1%	10.6	85.0	100	100
Series 5	5%	33.9	95.4	100	100
	10%	48.5	97.7	100	100
	1%	14.2	84.8	100	100
Series 6	5%	34.5	94.8	100	100
	10%	48.7	98.0	100	100
	1%	17.1	92.9	100	100
Series 7	5%	46.6	98.4	100	100
	10%	58.4	99.3	100	100
	1%	12.6	79.8	100	100
Series 8	5%	36.0	93.1	100	100
	10%	52.1	96.5	100	100
	1%	14.0	74.8	100	100
Series 9	5%	35.9	92.0	100	100
	10%	50.8	95.5	100	100

Note: Table 2 contains rejection frequencies of the null hypothesis of no change in the mean. Rejection frequencies are based on 1000 replications generated from the *Series* 4-9 where the nominal significance levels are 1% , 5% and 10%, the sample sizes are $n = 30, n = 100, n = 500$ and $n = 1000$.

From Table 2, we observe that, except for the small sample size n=30, the test has a good power either for homoskedastic time series (*Series* 4 and 7) or heteroskedastic time series (*Series* 5,6,8 and 9). Rejection frequencies of the null in abrupt change (*Series* 4, 5 and 6) are a somewhat greater than the ones corresponding to a smooth change (*Series* 7,8 and 9).

5 Application to the Stock index data

We consider the daily returns of S&P 500 index, $r_t = \log P_t - \log P_{t-1}$, where P_t is the daily closing level of the index between January 3, 1950 and November 17, 2008. We test changes in the mean of the returns r_t and the absolute returns $y_t = |r_t|$.

For the time series $(y_t), 1 \le t \le n$, the test statistic is given by

$$\mathcal{B}_n = \frac{1}{\sqrt{n}\widehat{\sigma}} \max_{1 \le k \le n} \left| \sum_{t=1}^{k} y_t - k\widehat{\mu} \right|,$$

where $\widehat{\sigma}^2 = \sum_{t=1}^{n}(y_t - \widehat{\mu})^2/n$, $\widehat{\mu} = \sum_{t=1}^{n} y_t/n$ and the corresponding *p-value* given by $p - value = 1 - F_{B_\infty}(\mathcal{B}_n)$. The cumulative distribution function of B_∞ is given by (see Billingsley (1968))

$$F_{B_\infty}(z) = 1 + 2\sum_{k=1}^{\infty}(-1)^k \exp\{-2k^2 z^2\}.$$

Although the distribution function F_{B_∞} involves an infinite sum, the series is extremely rapidly converging. Usually a few terms suffice for very high accuracy. For example, (see Massey (1952)) the $90\%, 95\%$, and 99% quantiles are $1.225, 1.359$ and 1.628 respectively. Note that the quantiles are reached with a high accuracy using only 2 terms i.e.

$$1 + 2\sum_{k=1}^{2}(-1)^k \exp\{-2k^2(1.225)^2\} = 0.9005625,$$

$$1 + 2\sum_{k=1}^{2}(-1)^k \exp\{-2k^2(1.359)^2\} = 0.9502443$$

and

$$1 + 2\sum_{k=1}^{2}(-1)^k \exp\{-2k^2(1.628)^2\} = 0.9900245$$

Applying our test to r_t yields $p - value = 0.291$ and hence the null hypothesis of no change in the mean is not rejected.

To check if to the time series $y_t = |r_t|$ is affected by breaks in the mean, we apply our test to y_t to detect change in the mean. We obtain $p-value = 0$ for y_t, which strongly supports change in the mean of the absolute returns of S&P 500 index between January 4,1950 and November 17, 2008

9

Appendix. Proofs

To prove Theorem 1 we will establish first a functional central limit theorem for heteroskedastic time series. Such theorem is independent of interest. Let $D = D[0,1]$ be the space of random functions that are right-continuous and have left limits, endowed with the Skorohod topology. The weak convergence of a sequence of random elements X_n in D to a random element X in D will be denoted by $X_n \Longrightarrow X$.

Consider a standard Gaussian white noise (ε_t), i.e. $E(\varepsilon_t) = 0$ and $\text{var}(\varepsilon_t) = 1$. Let (σ_t) satisfying (6) and

$$W_n(\tau) = \frac{1}{\overline{\sigma}_2 \sqrt{n}} \sum_{t=1}^{[n\tau]} \sigma_t \varepsilon_t, \quad \tau \in [0,1]. \tag{20}$$

Many Functional central limit theorems were established for covariance stationary time series, see Boutahar (2008) and the references therein. Note that the process $(\sigma_t \varepsilon_t)$ is not covariance stationary and hence Davydov's (1970) results can't be applied to obtain the weak convergence of W_n in the Skorohod space.

There are two sufficient conditions to have $W_n \Longrightarrow W$ (see Billingsley (1968):

i) the finite-dimensional distributions of W_n converge to the finite-dimensional distributions of W,

ii) W_n is tight.

Theorem A1. *Assume that (ε_t) is a standard Gaussian white noise and (σ_t) satisfying (6). Then*
$$W_n \implies W \tag{21}$$
where W is standard Brownian motion.

Proof. To prove that the finite-dimensional distributions of W_n converge to those of W it is sufficient to show that for all integer $r \geq 1$, for all $0 \leq \tau_1 < ... < \tau_r \leq 1$ and for all $(\alpha_1, ..., \alpha_r)' \in \mathbb{R}^r$,

$$Z_n = \sum_{i=1}^{r} \alpha_i W_n(\tau_i) \quad \longrightarrow \quad Z = \sum_{i=1}^{r} \alpha_i W(\tau_i). \tag{22}$$

Since Z_n is Gaussian with zero mean, it is sufficient to prove that

$$\text{var}(Z_n) \rightarrow \text{var}(Z) = \sum_{1 \leq i,j \leq r} \alpha_i \, \alpha_j \min(\tau_i, \tau_j). \tag{23}$$

10

For all (τ_i, τ_j)

$$\text{cov}(W_n(\tau_i), W_n(\tau_j)) = \text{var}(W_n(\min(\tau_i, \tau_j))$$

$$= \frac{1}{n\overline{\sigma}_2^2} \sum_{t=1}^{[n\min(\tau_i,\tau_j)]} \sigma_t^2$$

$$\to \min(\tau_i, \tau_j) \text{ as } n \to \infty,$$

since $\text{var}(Z_n) = \sum_{1 \le i,j \le r} \alpha_i \, \alpha_j \text{cov}(W_n(\tau_i), W_n(\tau_j))$, the desired conclusion (23) holds.

To prove the tightness of W_n it suffices to show the following inequality [Billingsley (1968), Theorem 15.6]

$$E\left(|W_n(\tau) - W_n(\tau_1)|^\gamma \, |W_n(\tau_2) - W_n(\tau)|^\gamma\right) \le (F(\tau_2) - F(\tau_1))^\alpha \qquad (24)$$

for some $\gamma \ge 0$, $\alpha > 1$, and F is a nondecreasing continuous function on [0,1], where $0 < \tau_1 < \tau < \tau_2 < 1$.

We have

$$E\left(|W_n(\tau) - W_n(\tau_1)|^2 \, |W_n(\tau_2) - W_n(\tau)|^2\right) = \frac{1}{n\overline{\sigma}_2^2} \left(\sum_{t=[n\tau_1]+1}^{[n\tau]} \sigma_t^2\right) \left(\sum_{t=[n\tau]+1}^{[n\tau_2]} \sigma_t^2\right)$$

$$\le C(\tau - \tau_1)(\tau_2 - \tau)$$

for some constant $C > 0$

$$\le C \ (\tau_2 - \tau_1)^2/2.$$

Consequently (24) holds with $\gamma = \alpha = 2$ and $F(t) = \sqrt{C/2}\, t$.

A1. Proof of Theorem 1

We have

$$B_n(\tau) = \frac{1}{\sqrt{n}\widehat{\sigma}} \sum_{t=1}^{[n\tau]} (y_t - \widehat{\mu})$$

$$= \left(\frac{\overline{\sigma}_2}{\widehat{\sigma}}\right) \frac{1}{\sqrt{n}\overline{\sigma}_2} \sum_{t=1}^{[n\tau]} \{(y_t - \mu) + (\mu - \widehat{\mu})\}$$

$$= \left(\frac{\overline{\sigma}_2}{\widehat{\sigma}}\right) \left\{W_n(\tau) - \frac{[n\tau]}{n} W_n(1)\right\}.$$

By using (21) it follows that

$$\left(\frac{\widehat{\sigma}}{\overline{\sigma}_2}\right) B_n \implies B,$$

11

and hence by continuous mapping theorem

$$\left(\frac{\widehat{\sigma}}{\overline{\sigma}_2}\right) \sup_{\tau \in [0,1]} |B_n(\tau)| \longrightarrow \sup_{\tau \in [0,1]} |B(\tau)|.$$

To achieve the proof of (7) it's sufficient to prove that

$$\widehat{\sigma} \xrightarrow{\quad P \quad} \overline{\sigma}_2. \tag{25}$$

Let $\mathcal{F}_t = \sigma - field\ (\varepsilon_1, ..., \varepsilon_t)$ and $F = (\mathcal{F}_n)$ the corresponding filtration. Then $N_n = \sum_{t=1}^{n} \sigma_t \varepsilon_t$ is a square integrable martingale adapted to F, with increasing process $\langle N_n \rangle = \sum_{t=1}^{n} \sigma_t^2$.

By using (6), $\langle N_n \rangle$ satisfies

$$\frac{\langle N_n \rangle}{n} \to \overline{\sigma}_2^2,$$

therefore (see Duflo (1997), theorem 1.3.15.)

$$\frac{1}{n} \sum_{t=1}^{n} \sigma_t \varepsilon_t \xrightarrow{\quad a.s. \quad} 0, \tag{26}$$

where $\xrightarrow{\ a.s.\ }$ denotes the almost sure convergence.

Likewise $M_n = \sum_{t=1}^{n} \sigma_t^2 (\varepsilon_t^2 - 1)$ is a square integrable martingale adapted to F, with increasing process $\langle M_n \rangle = 2 \sum_{t=1}^{n} \sigma_t^4$, hence theorem 1.3.15. in Duflo (1997) implies that

$$\frac{1}{\langle M_n \rangle} \sum_{t=1}^{n} \sigma_t^2 (\varepsilon_t^2 - 1) \xrightarrow{\quad} 0 \text{ almost surely on } \{\langle M_\infty \rangle = \infty\} \tag{27}$$

where $\langle M_\infty \rangle = \lim_{n \to \infty} \langle M_n \rangle$. Since

$$\left(\sum_{t=1}^{n} \sigma_t^2\right)^2 \leq n \sum_{t=1}^{n} \sigma_t^4, \tag{28}$$

The assumption (6) implies that there exist an universal constants $0 < K_1 < K_2 < \infty$ such that

$$K_1 < \frac{1}{n} \sum_{t=1}^{n} \sigma_t^2 < K_2,$$

this together with (28) implies that $\langle M_n \rangle \geq 2nK_1^2$ which implies that

$$\{\langle M_\infty \rangle = \infty\} = \Omega$$

12

and hence

$$\frac{1}{\langle M_n \rangle} \sum_{t=1}^{n} \sigma_t^2 (\varepsilon_t^2 - 1) \xrightarrow{a.s.} 0, \tag{29}$$

Since (σ_t) is a bounded deterministic sequence, then there exists an universal $K > 0$ such that $\sigma_t^4 \leq K$ for all $t \geq 1$, hence $\langle M_n \rangle \leq nK$ for all n, therefore

$$\left| \frac{1}{n} \sum_{t=1}^{n} \sigma_t^2 (\varepsilon_t^2 - 1) \right| = \frac{\langle M_n \rangle}{n} \left| \frac{1}{\langle M_n \rangle} \sum_{t=1}^{n} \sigma_t^2 (\varepsilon_t^2 - 1) \right|$$

$$\leq K \left| \frac{1}{\langle M_n \rangle} \sum_{t=1}^{n} \sigma_t^2 (\varepsilon_t^2 - 1) \right|,$$

using (29), it follows that

$$\frac{1}{n} \sum_{t=1}^{n} \sigma_t^2 (\varepsilon_t^2 - 1) \xrightarrow{a.s.} 0, \tag{30}$$

By using (26)-(30),

$$\frac{1}{n} \sum_{t=1}^{n} y_t = \mu + \frac{1}{n} \sum_{t=1}^{n} \sigma_t \varepsilon_t \tag{31}$$

$$\xrightarrow{a.s.} \mu$$

and

$$\frac{1}{n} \sum_{t=1}^{n} y_t^2 = \mu^2 + 2\mu \frac{1}{n} \sum_{t=1}^{n} \sigma_t \varepsilon_t + \frac{1}{n} \sum_{t=1}^{n} \sigma_t^2 + \frac{1}{n} \sum_{t=1}^{n} \sigma_t^2 (\varepsilon_t^2 - 1) \tag{32}$$

$$\xrightarrow{a.s.} \mu^2 + \overline{\sigma}_2^2$$

Combining (31) and (32) we obtain

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} y_t^2 - \left(\frac{1}{n} \sum_{t=1}^{n} y_t \right)^2$$

$$\xrightarrow{a.s.} \overline{\sigma}_2^2$$

and hence (25) follows.

A2. Proof of Theorem 2
For all $\tau < \tau_1$ we have

13

$$B_n(\tau) = B_n^0(\tau) + B_n^1(\tau) \tag{33}$$

where

$$B_n^0(\tau) = \frac{1}{\sqrt{n}\widehat{\sigma}} \sum_{t=1}^{[n\tau]} (\sigma_t \varepsilon_t - \widehat{\mu}_0), \widehat{\mu}_0 = \frac{1}{n} \sum_{t=1}^{n} \sigma_t \varepsilon_t \tag{34}$$

$$B_n^1(\tau) = \frac{1}{\sqrt{n}\widehat{\sigma}} \left\{ [n\tau]\mu_{(1)} - \frac{[n\tau]}{n} \left([n\tau_1]\mu_{(1)} + (n - [n\tau_1] + 1)\mu_{(2)} \right) \right\}.$$

Straightforward computation leads to

$$\widehat{\sigma}^2 \xrightarrow{a.s.} \sigma_*^2 = \overline{\sigma}_2^2 + \tau_1(1 - \tau_1)(\mu_{(1)} - \mu_{(2)})^2.$$

Therefore

$$B_n^0(\tau) \xrightarrow{\frac{\overline{\sigma}_2}{\sigma_*}} B(\tau) \tag{35}$$

and

$$\frac{B_n^1(\tau)}{\sqrt{n}} \xrightarrow{a.s.} \frac{\tau(1 - \tau_1)(\mu_{(1)} - \mu_{(2)})}{\sigma_*}.$$

Hence

$$B_n^1(\tau) \xrightarrow{a.s.} sgn(\mu_{(1)} - \mu_{(2)})\infty \tag{36}$$

where $sgn(x) = 1$ if $x > 0$ and -1 otherwise. Finally (33), (35) and (36) imply that

$$|B_n(\tau)| \xrightarrow{P} +\infty$$

and then the desired conclusion (9) holds.

A3. Proof of Theorem 3

$$B_n(\tau) = B_n^0(\tau) + B_n^1(\tau) \tag{37}$$

where $B_n^0(\tau)$ is given by (34) and

$$B_n^1(\tau) = \frac{1}{\sqrt{n}\widehat{\sigma}} \left\{ \sum_{t=1}^{[n\tau]} \mu_t - \frac{[n\tau]}{n} \sum_{t=1}^{n} \mu_t \right\}$$

$$= \frac{(\mu_{(2)} - \mu_{(1)})}{\sqrt{n}\widehat{\sigma}} \left\{ \sum_{t=1}^{[n\tau]} F(t/n, \tau_1, \gamma) - \frac{[n\tau]}{n} \sum_{t=1}^{n} F(t/n, \tau_1, \gamma) \right\}.$$

Straightforward computation leads to

14

$$\widehat{\sigma}^2 \xrightarrow{a.s.} \sigma_*^2 = \overline{\sigma}_2^2 + (\mu_{(2)} - \mu_{(1)})^2 \left\{ \int_0^1 F^2(x, \tau_1, \gamma) dx - \left(\int_0^1 F(x, \tau_1, \gamma) dx \right)^2 \right\}.$$

Therefore for all $\tau \in (0, 1)$

$$\frac{B_n^1(\tau)}{\sqrt{n}} \xrightarrow{a.s.} \frac{(\mu_{(2)} - \mu_{(1)})}{\sigma_*} T(\tau),$$

where

$$T(\tau) = \int_0^\tau F(x, \tau_1, \gamma) dx - \tau \int_0^1 F(x, \tau_1, \gamma) dx. \tag{38}$$

Moreover, there exists $\tau^* \in (0, 1)$ such that $T(\tau^*) \neq 0$, since if we assume that $T(\tau) = 0$ for all $\tau \in (0, 1)$ then

$$\begin{aligned} \frac{dT(\tau)}{d\tau} &= F(\tau, \tau_1, \gamma) - \int_0^1 F(x, \tau_1, \gamma) dx \\ &= 0 \end{aligned}$$

for all $\tau \in (0, 1)$ which implies that $F(\tau, \tau_1, \gamma) = \int_0^1 F(x, \tau_1, \gamma) dx = C$ for all $\tau \in (0, 1)$ or

$$\begin{aligned} \mu_t &= \mu_{(1)} + (\mu_{(2)} - \mu_{(1)})C \\ &= \mu \text{ for all } t \geq 1 \end{aligned}$$

and this contradicts the alternative hypothesis H_1.

$$\frac{B_n^1(\tau^*)}{\sqrt{n}} \xrightarrow{a.s.} \frac{(\mu_{(2)} - \mu_{(1)})}{\sigma_*} T(\tau^*)$$

and $T(\tau^*) \neq 0$ imply that

$$|B_n(\tau^*)| \xrightarrow{P} +\infty,$$

consequently, the desired conclusion (13) holds.

Remark 2. For the logistic transition, the function $T(\tau)$ in (38) is given by

$$\begin{aligned} T(\tau) &= \frac{1}{\gamma} \left\{ \tau \log \left[(1 + \exp(\gamma(\tau_1 - 1)))(1 + \exp(\gamma\tau_1)) \right] \right\} \\ &\quad - \frac{1}{\gamma} \left\{ \log \left[(1 + \exp(\gamma(\tau_1 - \tau)))1 + \exp(\gamma\tau_1)) \right] \right\}, \end{aligned}$$

15

and for the exponential transition

$$T(\tau) = \sqrt{\frac{\pi}{4\gamma}}\left\{(\tau - 1)erf(\sqrt{\gamma}\tau_1) + erf(\sqrt{\gamma}(\tau_1 - \tau)) - \tau erf(\sqrt{\gamma}(\tau_1 - \tau))\right\},$$

where erf is the Error function given by

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt.$$

References

[1] Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61**, 821– 856.

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

[3] Bai, J. and Perron, P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica*, **66,** 47–78.

[4] Billingslley, P. (1968) *Convergence of probability measures*. New York: Wiley.

[5] Boutahar, M. (2008) Identification of Persistent Cycles in Non-Gaussian Long Memory Time Series. *J. Time Ser. Anal.*, **Vol. 29**, Issue 4, 653-672.

[6] Chow, G.C. (1960) Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28,** 591-605.

[7] Davydov, Y.A. (1970) The invariance principle for stationary processes. *Theory Probab. Appl.*, **15**, 487-498.

[8] Ding, Z., Granger C.W.J. and Engle, R.F.(1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, **1**, 83-106.

[9] Duflo, M. (1997) *Random Iterative Models.* Springer-Verlag Berlin Heidelberg.

[10] Gombay, E. (2008) Change detection in autoregressive time series. *J. Multivariate Anal.*, **99**, 451-464.

[11] Hawkins, D.M. (1977) Testing a sequence of observations for a shift in location. *J. Amer. Statist. Assoc.*, **72**, 180–186.

[12] James, B., James, K. and Siegmund, D. (1987) Tests for a change point. *Biometrika*, **74**, 71– 83.

[13] Liu, J., Wu, S. and Zidek, J.V. (1997) On segmented multivariate regressions. *Statistica Sinica*, **7**, 497-525.

[14] Massey, F. J. (1952) Distribution Table for the Deviation between Two Sample Cumulatives. *Annals of Mathematical Statistics*, **23**, 435–441.

[15] Qu, Z. and Perron, P. (2007) Estimating and testing structural changes in multivariate regressions. Econometrica, **75**, 459–502.

[16] Quandt, R.E. (1960) Tests of the hypothesis that a linear regression obeys two separate regimes. *J. Amer. Statist. Assoc.*, **55**, 324–330.

[17] R Development Core Team, R: A language and environment for statistical computing. R, Foundation for Statistical Computing, Vienna, Austria. 2004. URL: (http://www.R-project.org)

[18] Sen, A. and Srivastava, M.S. (1975a) On tests for detecting change in mean. *Ann. of Statist.* **3,1**, 98–108.

[19] Sen, A. and Srivastava, M.S. (1975b) Some one-sided tests for change in level. *Technometrics*, **17**, 61– 64.

[20] Srivastava, M.S. and Worsley, K.J. (1986) Likelihood ratio tests for a change in the multivariate normal mean. *J. Amer. Statist. Assoc.*, **81**, 199–204.

[21] Starica, C. and Granger,C.W.J. (2005). Nonstationarities in Stock Returns. *The Review of Economics and Statistics*, **87(3)**, 503-522.

[22] Teräsvirta, T. (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. *J. Amer. Statist. Assoc.,* **89**, 208–218.

[23] Worsley, K.J. (1979) On the likelihood ratio test for a shift in locations of normal population. J*. Amer. Statist. Assoc.*, **74**, 365– 367.

[24] Yao, Y.C. (1988) Estimating the number of change-points via Schwarz' criterion. *Statistics & Probability Letters,* **6**, 181– 189.

[25] Yao, Y.C. and Au, S.T. (1989) Least squares estimation of a step function. *Sankhya*, **Series A 51**, 370–381.

[26] Yin, Y.Q. (1988) Detection of the number, locations and magnitudes of jumps. *Communications in Statistics Stochastic Models,* **4**, 445– 455.