Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Empirical Finance 11 (2004) 399–421
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/econbase

# Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns

Clive W.J. Granger[a,*], Namwon Hyung[b]

[a] *Department of Economics, Uinversity of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA*
[b] *Department of Economics, The University of Seoul, 90 Cheonnong-dong, Dongdaemoon-gu, Seoul, 130-743, South Korea*

Accepted 20 March 2003

**Abstract**

This paper shows that occasional breaks generate slowly decaying autocorrelations and other properties of $I(d)$ processes, where $d$ can be a fraction. Some theory and simulation results show that it is not easy to distinguish between the long memory property from the occasional-break process and the one from the $I(d)$ process. We compare two time series models, an occasional-break model and an $I(d)$ model to analyze S&P 500 absolute stock returns. An occasional-break model performs marginally better than an $I(d)$ model in terms of in-sample fitting. In general, we found that an occasional-break model provides less competitive forecasts, but not significantly. However, the empirical results suggest a possibility such that, at least, part of the long memory may be caused by the presence of neglected breaks in the series. We show that the forecasts by an occasional break model incorporate incremental information regrading future volatility beyond that found in $I(d)$ model. The findings enable improvements of volatility prediction by combining $I(d)$ model and occasional-break model.


*JEL classification:* C22; C50; G10
*Keywords:* Occasional structural breaks; Long memory; Autocorrelation; Absolute stock return

## 1. Introduction

There have been several articles analyzing the long-run properties of stock returns. Granger and Ding (1995a,b) considered long return series, using the well-known Standard

* Corresponding author. Tel.: +1-858-534-3383; fax: +1-858-534-7040.
*E-mail address:* cgranger@ucsd.edu (C.W.J. Granger).

0927-5398/$ - see front matter 
doi:10.1016/j.jempfin.2003.03.001

and Poor's (S&P) 500 index of about 17,000 daily observations, and established a set of temporal and distributional properties for such series. They suggested that the absolute returns are well characterized by an $I(d)$ process, but the parameter estimates of the $I(d)$ model sometimes vary considerably from one subseries to the next as shown by Granger and Ding (1996).

There have been several unsuccessful attempts to explain the cause of long memory and the observed time-varying property of $d$ by occasional regime switches. For example, Rydén et al. (1998) suggest that the temporal higher-order dependence observed in return series may be well described by a hidden Markov model. Such a model is estimated for the series considered by Granger and Ding (1995a,b); however, they failed to explain the one stylized fact which is the very slowly decaying autocorrelation function for absolute returns. On the other hand, Lobato and Savin (1998) investigate if the observed evidence of long memory is, in fact, due to nonstationarity during the long period. They split their sample (1962–1994) into two periods, taking January 1973 as the breaking point. But they do not find any evidence that long memory was caused by the structural break of 1973.

We, however, suspect that there are structural changes in the series since Granger and Ding (1996) examined very long time series of absolute stock returns from 1928 to 1991. If such structural changes exist, a stationary process that encounters occasional regime switches will have long memory properties that are similar to those of the $I(d)$ process, as pointed out by Granger (1998, comment). One plausible model to explain the (time-varying) long-memory property in the stock market might be derived from the example of Granger and Teräsvirta (1999), a nonlinear model with level changes.

In this paper, we provide some theoretical arguments and simulation results which show the claim that it is troublesome in practice to distinguish between the occasional breaks process and the $I(d)$ process. The plan of the paper is as follows. Section 2 introduces an $I(d)$ model briefly. Section 3 contains our analysis of long memory properties of a simple linear model with occasional, or infrequent, breaks in mean. In Section 4, we discuss spurious breaks in the $I(d)$ processes and over-difference phenomena caused by removing the estimated breaks. Section 5 is devoted to applications using the S&P 500 stock index and, finally, Section 6 presents conclusions.

## 2. *I(d)* process and long memory

Long memory has been used to model the persistence of stationary economic data ever since the work of Granger and Joyeux (1980). The $I(d)$ model is widely used for series with long memory, that is

$$(1-L)^d y_t = \varepsilon_t, \tag{1}$$

where $\varepsilon_t$ is zero mean, i.i.d. with variance $\sigma_\varepsilon^2$, for example. For any real $d > -1$, the fractional difference operator, $(1-L)^d$, is defined by its Maclaurin series, that is,

$$(1-L)^d = \sum_{j=0}^{\infty} \pi_j L^j,\ \pi_j = \frac{j-1-d}{j}\pi_{j-1},\ \pi_0 = 1. \tag{2}$$

From this expression, we can infer already a potential slow decay in the autocorrelation function. Traditionally, long memory has been defined in the time domain in terms of decay rates of the autocorrelations at high lags, that is, the long-term persistence. The long-term decay is solely determined by $d$, which is also called the memory parameter. There are several definitions of the property of 'long memory' for $-1/2<d<1/2$ (refer to Baillie, 1996). One definition is,

**Definition 1.** A time series, $\boldsymbol{y}_t$, possesses long memory, if the quantity

$$\lim_{T\to\infty}\sum_{k=-T}^{T}|\rho_k|$$

is non-finite, where $\rho_k$ is autocorrelation function of $y_t$.

Equivalently, Definition 1 implies that the spectral density $f(\omega)$ of this series is unbounded at low frequencies, i.e.,

$$f(\omega)\to\infty$$

as $\omega\to 0$. Alternatively, long memory can be defined in terms of decay rates of autocorrelations. An autocovariance function, $\gamma_k$, of the long memory process is given by

$$\gamma_k\approx c_k k^{2d-1}$$

where $d>0$ and $c_k$ is a slowly varying function at infinity. Then, the autocorrelation function $\rho_k$ is proportional to $k^{2d-1}$ as $k$ increases. Consequently, the autocorrelations of this process decay hyperbolically to zero as $k\to\infty$, in contrast to the exponential decay in a covariance-stationary ARMA process. Among the several definitions of long memory, we shall use Definition 1 for our analysis in Section 3.

There are several methods of testing long memory of time series. We use the following two methods in this paper. The first method of estimating the value of $d$ is the Geweke and Porter-Hudak (1983) (henceforth GPH) method, which is based on the following regression,[1]

$$\ln\{I(\omega_j)\}=c-d\ln\{4\sin^2(\omega_j/2)\}+u_j,\qquad j=1,\ldots,n. \tag{3}$$

where $I(\omega_j)=\frac{1}{2\pi}\left|\sum_{i=1}^{T}y_t\exp(i\omega_j t)\right|^2$ is the periodogram at frequency $\omega_j=2\pi j/T$. The window size $n$ depends on the sample size $T$. Then, the OLS estimator of $d$ will be asymptotically normal and with the variance $\pi^2/6n$. Alternatively, we use Lobato and Robinson's (1998) LM statistic, which tests $H_0$: $d=0$ against $H_a$: $d\neq 0$ (or $d>0$). Lobato and Robinson (1998) provide conditions that establish under the null that LM statistics converge to $\chi_1^2$ in distribution. For the detailed discussion and summary of the earlier work on $I(d)$ model, see Baillie's (1996) paper.

[1] As pointed out by a referee, Robinson (1995) corrected the log periodogram regression; $j$ cannot start at 1 in his regression. In our paper, however, we use the GPH method, because our results will not be affected by the choice of starting value of $j$.

## 3. Occasional break process and long memory

A time series may perhaps show long memory, not because it is really $I(d)$ but because of the neglected occasional breaks in the series. In this section, we show that occasional level shifts in mean give rise to the observed long memory phenomenon. The condition under which occasional level shifts in mean can exhibit long memory is described. Some comparison between this model and an $I(d)$ model are also made. For notational convenience, let an occasional-break model denote a time series model with occasional structural breaks in mean.

### *3.1. Occasional break process*

We shall consider a model with only a few breaks in the mean. For an example of this process, we can use an 'occasional-break model' (Chen and Tiao, 1990; Engle and Smith, 1999) as the following:

$$y_t = m_t + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{4}$$

where $\varepsilon_t$ is a noise variable and occasional level shifts, $m_t$, are controlled by two variables $q_t$ (date of break) and $\eta_t$ (size of jump), as

$$m_t = m_{t-1} + q_t \eta_t, \tag{5}$$

where $\eta_t$ is i.i.d. $(0,\sigma_\eta^2)$. For the simulation in the following sections, the distribution of $\eta_t$ is taken to be $N(0,\sigma_\eta^2)$ although this distribution has no particular relevance. We assume that $q_t$ follows an i.i.d. binomial distribution, that is,

$$q_t = \begin{cases} 0, \text{with probability } 1-p \\ 1, \text{with probability } p \end{cases} \tag{6}$$

The following assumption is used in this model throughout the paper.

**Assumption 1.** The probability of breaks $p$ converges to zero slowly as the sample size increases, i.e., $p \to 0$ as $T \to \infty$, yet $\lim_{T\to\infty} Tp$ is a non-zero finite constant.

This assumption implies that the expected number of breaks, $Tp$, is bounded from above even in the extreme case that $T$ increases to infinity. To preserve memory properties which are induced by breaks, the number of breaks has to remain finite as the size of sample increases; however, we do not interpret this sample size-dependent probability as a truly time-varying parameter. Assumption 1 is an example of letting $p$ decrease with sample size, so that regardless of the sample size, realization tends to have just finite breaks. For a related discussion, refer to Diebold and Inoue (2001).

Combine Eqs. (4) with (5), then $y_t$ is represented by

$$y_t = \{m_0 + q_1\eta_1 + \cdots + q_t\eta_t\} + \varepsilon_t. \tag{7}$$

The time-varying mean of $y_t$ is $\{m_0 + q_1\eta_1 + \ldots + q_t\eta_t\}$, which shows infrequent level shifts.

One problem with the binomial model in Eq. (6) is that this model implies sudden changes only. It may, however, be the case that structural changes occur gradually. For this reason, one may wish to use a regime switching model, which is actually a simple generalization of Eq. (6). Let $s_t$ be a latent random variable with two discrete values: 0,1. Each value of $s_t$ represents a different state in the length of memory of shock. $s_t$ is assumed to be governed by the following Markov probability law: $p_{ij}=\Pr(s_t=j|s_{t-1}=i)$. Then, one can use a regime switching model of $q_t$ such that $q_t=0$, when $s_t=0$, and $q_t=1$, when $s_t=1$. In this specification, the state of $s_t$ will determine if the shock of $\eta_t$ is permanent or not. A regime with $s_t=1$ represents a period of structural change.

Graphical examples of these DGPs are provided in Fig. 1, which will be explained in detail in Section 3.4.

### *3.2. The sample autocorrelation function*

Now, we examine the sample autocorrelation of occasional break model when the probability of level shift is small. We consider time series of Eq. (7) for a finite number of observations, $T$, which allows us to focus on the effects of occasional level shifts on memory.

For simplicity, let $\varepsilon_s$, $\eta_t$ and $q_\tau$ be independent for all $s$, $t$ and $\tau$. For the initial conditions, assume $m_0=0$, $q_t=0$, $\varepsilon_t=0$ and $\eta_t=0$ for all $t\leq 0$, then the mean of $y_T$ is $E(y_T)=0$, and its variance is

$$\mathrm{var}(y_T) = Tp\sigma_\eta^2 + \sigma_\varepsilon^2$$

Similarly, one finds that the covariance between $y_T$ and $y_{T+k}$ is

$$\mathrm{cov}(y_T, y_{T+k}) = Tp\sigma_\eta^2$$

Taken together, one gets the following $k$-th autocorrelation equation of this process,

$$\mathrm{corr}(y_T, y_{T+k}) = \frac{Tp\sigma_\eta^2}{\sqrt{Tp\sigma_\eta^2 + \sigma_\varepsilon^2}\sqrt{(T+k)p\sigma_\eta^2 + \sigma_\varepsilon^2}}$$

It is obvious that the sum of absolute values of autocorrelations are not bounded as $T\rightarrow\infty$, even with finite number of breaks.

**Proposition 1.** *A time series is generated from Eqs. (4–6), and if the probability of breaks p satisfies Assumption 1, then an occasional-break model possesses a long memory by Definition 1.*

The proof of this proposition is obvious. $\mathrm{corr}(y_T, y_{T+k}) \rightarrow (1+\frac{\sigma_\varepsilon^2}{Tp\sigma_\eta^2})^{-1} > 0$ as $T\rightarrow\infty$ for all $k$.

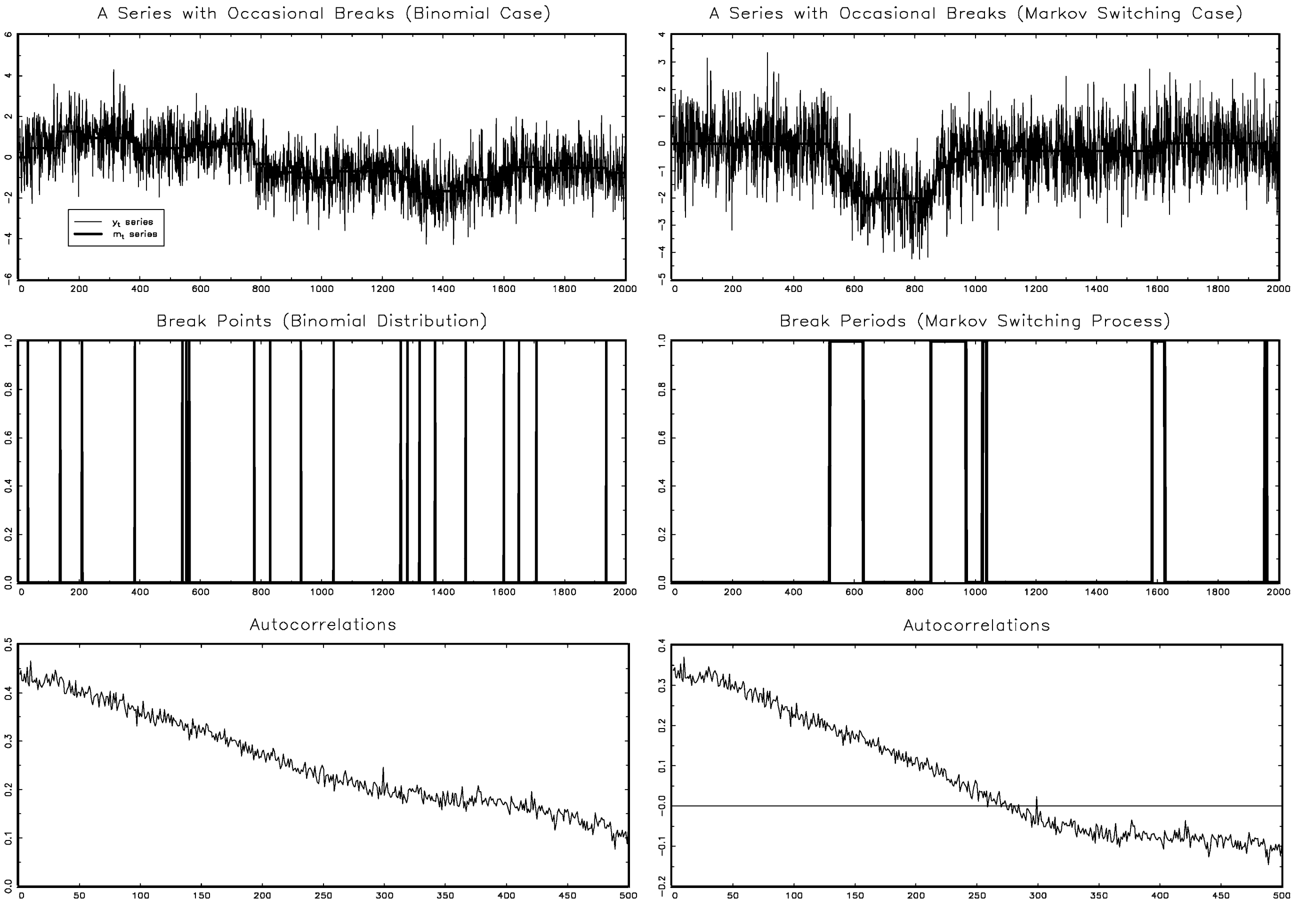


Fig. 1. (1) A series with occasional breaks (binomial case). (2) A series with occasional breaks (Marcov switching case).

We can derive properties of a sample autocorrelation function. Our subsequent results are based on the sample autocorrelation function. After some tedious algebra, we get the following approximation of the sample autocorrelation function,

$$\hat{\rho}_T(k) = \frac{\sum_{t=1}^{T-k}(y_t-\bar{y})(y_{t+k}-\bar{y})}{\sum_{t=1}^{T}(y_t-\bar{y})^2} \approx \frac{\frac{Tp\sigma_\eta^2}{6}\left(1-\frac{k}{T}\right)\left(1-2\frac{k}{T}+4\left(\frac{k}{T}\right)^2\right)}{\frac{Tp\sigma_\eta^2}{6}+\sigma_\varepsilon^2} \tag{8}$$

where $\bar{y}=\frac{1}{T}\sum_{t=1}^{T}y_t$ and $\approx$ denotes approximate equality for any $k$ such that $k/T\rightarrow 0$ as $T$ increases. The value of $Tp$ implies an expected number of structural breaks with the sample size $T$, and $\sigma_\eta^2$ is related to the size of breaks. As explained in Proposition 2, these parameters are closely linked to the memory properties of the occasional break model.

**Proposition 2.** *A time series is generated from Eqs. (4–6), and if the probability of breaks p satisfies Assumption 1, then the k-th sample autocorrelation in Eq. (8), $\rho_T(k)$, converges to nonzero value for any k such as $k/T\rightarrow 0$ as $T\rightarrow\infty$,*

$$\hat{\rho}_T(k)\rightarrow\left(1+\frac{6\sigma_\varepsilon^2}{Tp\sigma_\eta^2}\right)^{-1}$$

*where* $0<\left(1+\frac{6\sigma_\varepsilon^2}{Tp\sigma_\eta^2}\right)^{-1}<1$.

**Remark 1.** There are two trivial cases. First, if $0<p<1$ and p is fixed, the break process is obviously I(1) and $\hat{\rho}_T(k)\rightarrow 1$ as $T\rightarrow\infty$. Second, when $p=0$, i.e., no break, then this process is I(0) and $\hat{\rho}_T(k)\rightarrow 0$ as $T\rightarrow\infty$.

We focus on the case of long-memory caused by occasional breaks in this paper. Proposition 1 shows that a series with finite breaks may have similar autocorrelations to an $I(d)$ process. Typically the $I(d)$ process has slow hyperbolically decaying autocorrelations after the initial drop-off from $k=0$ to $k=1$. The autocorrelations in Eq. (8) do not decline exponentially (even if $\varepsilon_t$ of Eq. (4) has serial correlation), but decays very slowly as $k$ increases. For any given $k$, the sample autocorrelation approaches a nonzero constant as $T\rightarrow\infty$. As $Tp$ increases, there are more breaks and the value of the sample autocorrelation tends to be higher. An increase of $\sigma_\eta^2$, which means a larger magnitude of breaks, has similar effects on the autocorrelations. The intuition is that increases of $Tp$ or $\sigma_\eta^2$ make an occasional-break process closer to a random walk.

There are other properties of the $I(d)$ process which are heavily used by researchers, which are also true for a series with breaks. Granger and Marmol (1997) show that the correlogram of $I(d)$ is low, but remains positive for many lags. We found that it is also true for a series with breaks from Proposition 2 and from unreported simulation studies. Mikosch and Stărică (1999) show similar results from the volatility series with breaks. These findings show the long memory property of the occasional-break process in a finite sample. In Section 3.3, we show the long memory property of an occasional break process in the frequency domain.

### 3.3. Long memory in the frequency domain

In Section 3.2, we showed that neglected breaks generate a long memory effect in the autocorrelation function. In this section, we present similar analysis in the frequency domain. We show the bias of the GPH estimator explicitly when there are neglected breaks.

Suppose that $m_t$ is the series from Eq. (5). From filtering considerations, the spectrum of $m_t$ can be thought of as

$$f_m(\omega) = |1 - z|^{-2}\frac{1}{2\pi}p\sigma_\eta^2, \omega \neq 0$$

where $z = e^{-i\omega}$. For small $\omega$,

$$f_m(\omega) = c\omega^{-2}$$

where $c = p\sigma_\eta^2/2\pi$. Since $m_t$ is independent of $\varepsilon_s$ for all $t$ and $s$, the spectrum of $y_t$ in Eq. (4) can be thought of as

$$f_y(\omega) = f_m(\omega) + f_\varepsilon(\omega) = |1 - z|^{-2}\frac{1}{2\pi}p\sigma_\eta^2 + \frac{1}{2\pi}\sigma_\varepsilon^2, \omega \neq 0 \tag{9}$$

It follows that

$$f_y(\omega) = c(\omega)^{-2} + c', \tag{10}$$

for $\omega$ small, where $c = p\sigma_\eta^2/2\pi$, $c' = \sigma_\varepsilon^2/2\pi$.

When attention is confined to frequencies near zero, the differencing parameter can be estimated consistently from the GPH regression since

$$-\frac{1}{2}\frac{\partial \ln f_z(\omega)}{\partial \ln\omega} = d \tag{11}$$

with a time series $z_t$ which is $I(d)$ process where $f_z(\omega)$ is the spectrum of $z_t$. The relationship underlying Eq. (11) holds only for frequencies close to zero. For example, the consistency of the estimate $d$ in the GPH regression (3) requires that the window size $n$ grows with sample size, but at a slow rate, which ensures $\omega \to 0$.

Similarly, the elasticity in Eq. (10) is a negative, but a nonlinear function of $\omega$ such as

$$-\frac{1}{2}\frac{\partial \ln f_y(\omega)}{\partial \ln \omega} = \left(1+\omega^2\frac{\sigma_\varepsilon^2}{p\sigma_\eta^2}\right)^{-1}. \tag{12}$$

The RHS of Eq. (12) is increasing as $\omega \to 0$. It is evident that the estimated value of $d$ from GPH regression depends on the probability of breaks and on the relative variance of the innovations. For some $\omega$ close to zero frequency, but which approaches zero slowly such that $\omega^2/p \to \mathrm{C}$ as $\omega \to 0$, then the elasticity is

$$-\frac{1}{2}\frac{\partial \ln f_y(\omega)}{\partial \ln \omega} \to \left(1+C\frac{\sigma_\varepsilon^2}{\sigma_\eta^2}\right)^{-1}. \tag{13}$$

where $C$ is a non-zero, finite constant number. Thus, the time series $y_t$ seems to have long memory since the elasticity of spectrum near zero frequency is bound away from zero.

In the trivial case of no break, $p=0$, the elasticity is $-\frac{\partial \ln f_y(\omega)}{\partial \ln \omega}=0$ from Eq. (9), which implies no long memory in the frequency domain. Another obvious case $0<p<1$ and $p$ is fixed, then $-\frac{1}{2}\frac{\partial \ln f_y(\omega)}{\partial \ln \omega} \to 1$ as $\omega \to 0$, which is the unit root case.

The log-periodogram regression estimation procedure in Eq. (3) by Geweke and Porter-Hudak (1983) would suffer from a serious bias, if there are neglected breaks since the GPH estimation is focusing on the elasticity of the spectral density close to zero frequency.

**Proposition 3.** *When $y_t$ is generated from Eqs. (4–6), and the probability of breaks $p$ satisfies Assumption 1, then the estimated value of $d$ by the GPH method in Eq. (3) using $n=T^{1/2}$ would be non-zero positive. One would get a positive value of $d$ by the GPH only because $p$ is small enough (but not zero) to counter-balance small $\omega$, that is, $\omega \to 0$ but $\omega^2/p=O(1)$.*

The Proposition 3 is confirmed along the lines of simulation studies in the following section. A heuristic comment of this proposition follows: we can calculate a lower bound for the GPH estimator from Eq. (13) since the RHS function in Eq. (12) is monotonically increasing as $\omega \to 0$ and the GPH estimator is based on $\omega_j$, $j=1,\ldots,n$. For example, when $n=T^{1/2}$ then $C=\frac{4\pi^2}{Tp}$ at $\omega_n=2\pi n/T$. The lower bound of the RHS function in Eq. (12) is[2]

$$\left(1+4\pi^2\frac{\sigma_\varepsilon^2}{Tp\sigma_\eta^2}\right)^{-1}. \tag{14}$$

As $\frac{\sigma_\varepsilon^2}{Tp\sigma_\eta^2}$ decreases, the estimated value of $d$ is likely to increase, as well. This result holds for which all parameters are known and $T\to\infty$. It may not necessarily carry over to estimated processes in finite samples.

[2] The choice of $n=T^{1/2}$ is not the only necessary condition to obtain the Proposition 3. So long as $\omega^2/p=O(1)$ $p=O(1)$ as $\omega \to 0$, any choice of $n$ would provide a similar result with Proposition 3.

### 3.4. Simulation

In this section, we discuss some simulation studies. The model used in the simulation is the change in mean model from Eqs. (4) and (5). The following two artificial data sets illustrate the long memory property of occasional breaks model with different parameters. It should be noted that even though the series in the following examples are not $I(d)$ processes, the correlograms and the GPH estimators show a clear evidence of long memory.

**Example 1.** Let $q_t$ have a binomial distribution of Eq. (6). $T=2000$, $\sigma_\varepsilon^2=1$, $\sigma_\eta^2=0.25$, $p=0.01$. If $Tp$ is a small positive integer, a plot of $y_t$ against $t$ shows about 20 breaks in level since the value $Tp$ is an expected number of breaks within the sample period. The third graph of Fig. 1(1) plots sample autocorrelations of this series up to lag 500. Sample autocorrelations start around 0.44 (quite close to 0.45 calculated by Eq. (8) with $k=1$), but decrease very slowly. These figures might suggest that this series has long memory. By the GPH method, $\hat{d}$, is 0.747 with $t$-value 6.17, a clear indication of long memory. For the first differenced series, $\hat{d}$ is $-0.212$ with the standard error 0.13, which implies that taking difference of this series may cause over-difference.

**Example 2.** A series from Eqs. (4) and (5) when $q_t$ has a regime switching process instead of Eq. (6). $T=2000$, $\sigma_\varepsilon^2=1$, $\sigma_\eta^2=0.01$, $p_{11}=0.998$, $p_{22}=0.99$, implying unconditional probability of state 2 (break) is 0.167. The second graph of Fig. 1(2) plots the value of $q_t$, which is determined by the state variable, and for this series $\hat{d}$ is 0.860 with $t$-value 5.38. We observe gradual changes in the series when $q_t=1$. For the first differenced series, $\hat{d}$ is $-0.103$ with the standard error 0.11.

In Table 1, the number of replications for each experiment was 1000 with a sample size of 2000, $\sigma_\varepsilon^2=1$ and different values of $p$ and $\sigma_\eta^2$. In the first rows of Table 1, we report averaged values of the estimated sample autocorrelations. The numbers in brackets are

Table 1
Autocorrelation and GPH estimates of the occasional break process

| $p(Tp)$ | 0.0025(5) | 0.005(10) | 0.01(20) | 0.05(100) |
|---|---|---|---|---|
| $\sigma_\eta^2=0.005$ | 0.0046 [0.0041] | 0.0085 [0.0083] | 0.0163 [0.0164] | 0.0741 [0.0768] |
| | 0.0020 [0.0036] | 0.0051 [0.0071] | 0.0112 [0.0142] | 0.0549 [0.0665] |
| | 0.076 (0.70) | 0.119 (1.09) | 0.178 (1.64) | 0.384 (3.51) |
| 0.01 | 0.0086 [0.0083] | 0.0163 [0.0164] | 0.0314 [0.0322] | 0.1328 [0.1426] |
| | 0.0051 [0.0071] | 0.0110 [0.0142] | 0.0228 [0279] | 0.0989 [0.1235] |
| | 0.118 (1.08) | 0.175 (1.60) | 0.252 (2.31) | 0.488 (4.49) |
| 0.05 | 0.0379 [0.0399] | 0.0708 [0.0768] | 0.1293 [0.1426] | 0.3936 [0.4539] |
| | 0.0272 [0.0346] | 0.0524 [0.0665] | 0.0970 [0.1235] | 0.2873 [0.3930] |
| | 0.265 (2.41) | 0.363 (3.30) | 0.477 (4.40) | 0.736 (7.05) |
| 0.1 | 0.0699 [0.0768] | 0.1258 [0.1426] | 0.2174 [2496] | 0.5431 [0.6241] |
| | 0.0512 [0.0665] | 0.0936 [0.1235] | 0.1622 [0.2161] | 0.3903 [0.5403] |
| | 0.352 (3.22) | 0.464 (4.23) | 0.587 (5.47) | 0.825 (8.09) |

The values are averaged from 1000 simulations with 2000 sample size and $\sigma_\varepsilon^2=1$. $Tp$ means expected number of structural breaks and $\sigma_\eta^2$ is related to the size of breaks. The numbers in the first row are averaged values of the autocorrelations and the numbers in brackets are theoretical values from Eq. (8) with $k=1$. The numbers in the second row are the autocorrelation when $k=100$. The numbers in the third row are the estimated values of $d$ by the GPH method and $t$-values are in the parentheses.

theoretical values of autocorrelation in Eq. (8) when $k=1$. The second rows contain the values of autocorrelations when $k=100\%$, which decay little from the values when $k=1$. The average of estimated value $d$ by the GPH is presented in the third rows. When $\sigma_\eta^2=0.01$ as $Tp$ increases from 10 to 20, the averaged value of estimated $d$ increases from 0.175 to 0.252. As $Tp$ increases there are more breaks and a higher value of $\hat{d}$ is obtained from the GPH regression. An increase of $\sigma_\eta^2$ has similar effects on $\hat{d}$ since breaks are more likely to be detected. The percentiles $\hat{d}$ of the last column of Table 1 are presented in Fig. 2. For reference, the theoretical lower bounds for the GPH estimator in Eq. (14) are 0.0125, 0.0247, 0.1124 and 0.2021 for $\sigma_\eta^2=0.005$, 0.01, 0.05 and 0.1, respectively, with $Tp=100$, $\sigma_\varepsilon^2=1$. The majority of the GPH estimates are above these lower bounds, as expected. And we observe a clear tendency of increasing $d$ as $\sigma_\eta^2$ increases.

In Table 2, Lobato and Robinson's LM test was conducted for the various values of parameters, $T=200$, $p=0.025$, 0.05, 0.1, $\sigma_\varepsilon^2=1$, and $\sigma_\eta^2=0.001$, 0.005, 0.01, 0.05, 0.1, 0.5, 1. As the value of $Tp$ or $\sigma_\eta^2$ is getting larger, the rejection rate of the null hypothesis of no-long-memory is increasing spuriously. Overall, tests by LM statistics and the GPH method show similar results. Note that the rejection rates are similar for the same value of $Tp{\cdot}\sigma_\eta^2$. This confirms the conjecture of Section 3.3, that is, the value $Tp{\cdot}\sigma_\eta^2$ determines the degree of long memory of the occasional break model where $\sigma_\varepsilon^2$ remains equal for all cases. For example, when $p=0.1$ (20 expected breaks for sample size 200) and $\sigma_\eta^2=0.05$, the rejection rate of the LM test is 52.8%, and on the other hand the rejection rate is 54.4% when $p=0.05$ (10 breaks) and $\sigma_\eta^2=0.1$.

To sum up, we find that an occasional break model generates a long memory property. Similar to the examples of Granger and Teräsvirta (1999), an occasional break model has long-memory rather than short-memory, if we just consider the linear properties of the data, such as autocorrelation. So, disentangling of the properties of autocorrelation of break model and $I(d)$ model becomes difficult as the size of the break or the number of breaks increases. Furthermore, we get the time varying property of $d$ which Granger and Ding (1996) find in the financial time series. From other simulation studies (not reported here), we find that the values of estimated $d$ depend on the realized breaks, which are

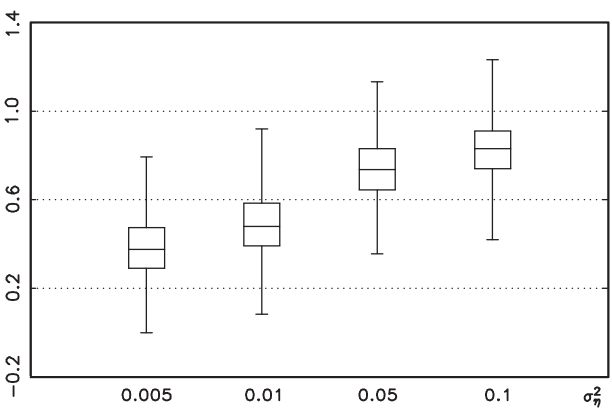


Fig. 2. Quantiles of the estimated d by GPH ($\sigma_\varepsilon^2=1$, $Tp=100$).

Table 2
Rejection rates of the null of stationarity against $I(d)$ process in the occasional break process

| $p(Tp)$ | 0.025(5) | 0.05(10) | 0.1(20) |
|---|---|---|---|
| $\sigma_\eta^2=0.001$ | 6.6 | 6.4 | 8.5 |
| | 1.0 | 1.6 | 1.9 |
| $\sigma_\eta^2=0.005$ | 7.2 | 9.9 | 12.9 |
| | 1.7 | 3.3 | 7.2 |
| $\sigma_\eta^2=0.01$ | 10.4 | 12.7 | 20.0 |
| | 2.5 | 7.0 | 14.0 |
| $\sigma_\eta^2=0.05$ | 23.6 | 36.5 | 55.0 |
| | 17.7 | 33.6 | 52.8 |
| $\sigma_\eta^2=0.1$ | 33.6 | 52.6 | 72.7 |
| | 30.9 | 54.4 | 70.6 |
| $\sigma_\eta^2=0.5$ | 69.2 | 86.1 | 93.1 |
| | 68.9 | 85.9 | 93.9 |
| $\sigma_\eta^2=1$ | 81.6 | 91.1 | 95.7 |
| | 80.2 | 90.0 | 94.8 |

The first numbers equal the % of $t$-value>1.645 in GPH. The numbers in the second row are the % of $p$-value of LM test which is less than 0.05. The results are based on 1000 replications with 200 sample size.

varying across subsamples when the series was generated by an occasional break model. A similar result was shown by Granger and Teräsvirta (1999).

## 4. *I(d)* process and spurious breaks

In the previous sections, we discussed that structural change is easily confused with the fractional integration. There exists another type of difficulty: the fractional integration of the DGP causes many breaks to be detected spuriously by standard estimation methods. Unlike $I(0)$ processes, $I(d)$ or $I(1)$ processes would have different effects on the estimated number of breaks. Under assumptions of weakly dependent, heterogenous error term, the consistency of a quasi-maximum likelihood estimator, or the least-squares estimator of the break, is well demonstrated by several researchers (Yao, 1988; Nunes et al., 1995). But, the estimator shows quite different properties for the time series that is integrated by fraction or integer. One might get a break spuriously near the middle of the time series even though there is no break. When the DGP is an integrated or fractionally integrated series without breaks, spuriously many breaks (but with different numbers of break depending on the value of $d$, as shown in the following simulation study) will be inferred.

Let us consider the Schwarz or Bayesian Information criterion as a method of choosing the number of breaks in mean. The $R$ breaks are given in Eq. (4) where $m_t=\mu_r$ for $r\in(k_{r-1},k_r]$, $r=1,\ldots,R+1$, such that $0=k_0<k_1<k_2<\ldots<k_R<k_{R+1}=T$. The estimator of $\sigma_\varepsilon^2$ is

$$\hat{\sigma}_\varepsilon^2(R)=\min_{0<k_1<\cdots<k_R<T}\frac{1}{T}\sum_{t=1}^{T}y_t^2-\frac{1}{T}\sum_{r=1}^{R+1}\frac{1}{k_r-k_{r-1}}\left(\sum_{t=k_{r-1}+1}^{k_r}y_t\right)^2. \tag{15}$$

Table 3
Percentage of breaks selected by the SBC

| $R_U$ | $R$ | $I(0)$ | $d=0.2$ | $d=0.4$ | $d=0.6$ | $d=0.8$ | $I(1)$ |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 95.8 | 59.9 | 22.0 | 5.5 | 1.0 | 0.1 |
| | 1 | 4.2 | 40.1 | 78.0 | 94.5 | 99.0 | 99.9 |
| 2 | 0 | 94.6 | 49.3 | 9.7 | 0.8 | 0.0 | 0.0 |
| | 1 | 4.1 | 26.3 | 24.1 | 10.6 | 3.0 | 0.5 |
| | 2 | 1.3 | 24.4 | 66.2 | 88.6 | 97.0 | 99.5 |
| 3 | 0 | 94.6 | 48.3 | 8.5 | 0.8 | 0.0 | 0.0 |
| | 1 | 4.0 | 24.4 | 15.5 | 3.7 | 0.6 | 0.0 |
| | 2 | 1.2 | 19.5 | 31.6 | 16.1 | 4.9 | 1.5 |
| | 3 | 0.2 | 7.8 | 44.4 | 79.4 | 94.5 | 98.5 |
| 4 | 0 | 94.6 | 48.2 | 7.9 | 0.8 | 0.0 | 0.0 |
| | 1 | 4.0 | 24.3 | 14.9 | 2.6 | 0.2 | 0.0 |
| | 2 | 1.2 | 18.2 | 24.6 | 8.8 | 1.1 | 0.1 |
| | 3 | 0.2 | 6.3 | 26.7 | 21.4 | 9.4 | 2.5 |
| | 4 | 0.0 | 3.0 | 25.9 | 66.4 | 89.3 | 97.4 |

Results based on 1000 replications with 100 observations. We consider six different DGPs: white noise, $I(d)$ processes and unit root process.

The estimated number of break points $R$ is found by

$$\arg\min_R \log(\hat{\sigma}_\varepsilon^2(R)) + (1+2R)\frac{\log T}{T} \tag{16}$$

subject to $R<R_U$, with $R_U$ a given fixed upper bound for $R$.

When $y_t$ is generated from the DGP of Eq. (1), with no break, one finds break points near the middle of the time series when $d>0$ (see Kuan and Hsu, 1998). Lavielle and Moulines (2000) show that, to get a consistent number of breaks in the least squares estimation, typically the penalization for the time series with $0<d<1/2$ should be higher than the one of Eq. (16), such as $R\log(T)/T^{1-2d}$. Otherwise, the estimated $R$ of Eq. (16) in the $I(d)$ process results in over-estimation of the number of breaks. This is true for $d\geq 1/2$.

Table 4
Estimated number of breaks and memory parameter in $I(d)$ process and residuals

| DGP | No. of breaks | $y_t$ | | $y_t-\hat{m}_t$ | |
|---|---|---|---|---|---|
| $d$ | | $\hat{d}$ | LM | $\hat{d}$ | LM |
| 0.2 | 2.91 | 0.203 (1.90) | 8.13 (0.12) | − 0.006 ( − 0.06) | 0.95 (0.51) |
| 0.4 | 8.35 | 0.410 (3.82) | 40.79 (0.00) | − 0.159 ( − 1.42) | 1.13 (0.44) |
| 0.6 | 13.36 | 0.620 (5.79) | 101.88 (0.00) | − 0.285 ( − 2.49) | 1.71 (0.32) |
| 0.8 | 17.10 | 0.827 (7.79) | 167.98 (0.00) | − 0.33 ( − 2.87) | 1.91 (0.28) |

Occasional breaks are estimated by Bai's (1997) method: To obtain consistent estimates for the error variance, we assume the number of breaks is 20 in the first step, where the size of test is chosen to be 0.05 with corresponding critical values 10.61. After identifying break dates, decompose $y_t$ into break component ($\hat{m}_t$) and break-free component (i.e., residuals $=y_t-\hat{m}_t$). $\hat{m}_t$ is a sample mean of $y_t$ of each regime. $d$ and LM are estimated by the GPH method and Lobato and Robinson method, respectively. $t$-statistics for $d$ and $p$-values for LM are given in parentheses below statistic values. Results are averaged based on 1000 replications with 2000 observations.

For the cases when $d \geq 1$, Bai (1998) provides rigorous proof. The estimated number of breaks is closely related to the value of $d$ as shown by the following proposition and the small sample simulation in Tables 3 and 4.

**Proposition 4.** *Suppose the DGP of $y_t$ is given by Eq. (1), with no break, then the expected value of estimated numbers of breaks, $R_T$, in Eq. (16) as $T \to \infty$ are (a) $R_T \to 0$, with $d=0$; (b) $R_T \to \bar{R} < \infty$, with $0<d<1/2$. For any choice of $R_U$, (c) $R_T \to R_U$, with $d>1/2$.*

**Proof.** (a) Refers the proof from Yao (1988) and Nunes et al. (1995). (b) The first part of Eq. (16) is bounded as

$$\hat{\sigma}_{\varepsilon}^2(R) \leq \frac{1}{T}\sum_{t=1}^{T} y_t^2 \to \mathrm{Var}(y_t)$$

in probability. Thus, the number of breaks cannot grow to infinity since the increase of $R$ will be penalized by the second term of Eq. (16). (c) The sums of squared residual for given $R$ breaks in Eq. (16) obeys

$$\frac{1}{T^{2d-1}}\hat{\sigma}_{\varepsilon}^2(R) = O_p(1). \tag{17}$$

Since the second term of Eq. (16) goes to zero as $T$ increase, Eq. (17) implies that, for any fixed $R$, only the first term in Eq. (16) matters asymptotically. Since $\hat{\sigma}_{\varepsilon}^2(R)$ is monotonically decreasing in $R$ for any given data set, an Eq. (16) is minimized at $R=R_U$ for $T$ large enough. Therefore, if one allows $T$ increases to infinity for any level of $R_U$, one tends to choose $R_U$ as an estimated number of breaks. □

In small sample, the estimator of the number of breaks will have some distribution. In Table 3, we present the simulation results of Proposition 3 in small samples. In our experiments, we generate six different DGPs. $I(d)$ series are generated by using Eq. (2), where it is assumed $\pi_j=0$ for $j>1000$ and the first 2000 observations are discarded. The numbers of breaks are estimated by Eq. (16). We set the maximum possible number of breaks, $R_U$, at 1, 2, 3 and 4. Table 3 shows a positive relation between the number of breaks and the value of $d$ in a finite sample. When the DGP is an $I(d)$ with $d>1/2$ or a random walk, the maximum permitted number of breaks are selected on the majority of occasions. Clearly, the $I(d)$ process is an intermediate process between $I(0)$ and $I(1)$ in terms of the estimated number of breaks in finite sample. Nunes et al. (1996) simulated the cases of white noise and a random walk. We find similar results in the first and last columns of Table 3.

Next, we estimate breaks by another method to show the robustness of our results. Further, we examine the effects of removing break components to the estimated value of $d$. In Table 4, the number of breaks is estimated by Bai's (1997) method[3] for various values of $d$ in the DGP. Fig. 3 shows examples of $I(d)$ processes and estimated means of each

[3] The admissible break points in the estimation procedure are given symmetrically with $\epsilon$, which indicates the fraction of break point $\frac{k}{T} \in [\epsilon, 1-\epsilon]$. We calculate critical values via 10,000 simulation with 10,000 sample size, when $\epsilon=0.01, 0.02$. For details, refer to Bai and Perron (1998). In each case, corresponding critical values are 10.61 and 10.26 with the size of test 0.05. The choice of $\epsilon$ is corresponding to the assumption of minimum observations between two consecutive breaks. When $T=2000$, $\epsilon=0.01$ implies that the minimum number would be 20, for example.

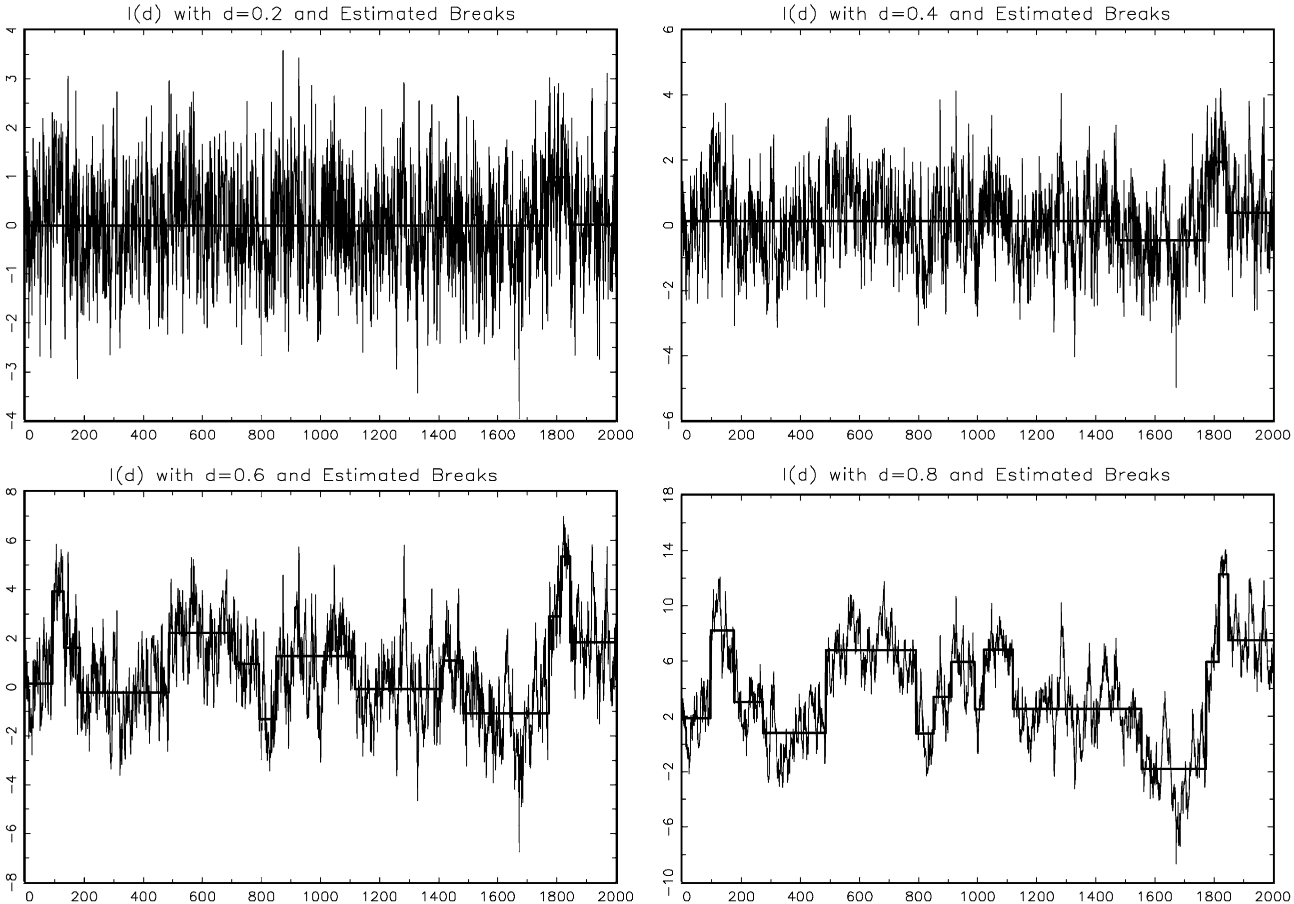


Fig. 3. (1) $I(d)$ with $d = 0.2$ and estimated breaks. (2) $I(d)$ with $d = 0.4$ and estimated breaks. (3) $I(d)$ with $d = 0.6$ and estimated breaks. (4) $I(d)$ with $d = 0.8$ and estimated breaks.

regime. In these specific realizations of $I(d)$ process, 2 breaks are detected when $d=0.2$, 3 breaks when $d=0.4$, 13 breaks when $d=0.6$ and 14 breaks when $d=0.8$. These graphs clearly suggest a positive relation between the number of breaks and the value of $d$ as Table 3. After removing breaks from the original series, we find some evidence of over-differencing, i.e., the estimated value of $d$ is less than zero in the fifth column of Table 4.

The break estimation methods reflect what would be inferred from visual inspection. It appears that such inspection is very dependent on DGP. The simulation results suggest that caution should be exercised in estimating breaks of fractionally integrated series. Indeed, as the degree of integration of DGP increases, more breaks are inferred in a finite sample. It reflects the facts that a fractionally integrated process itself contains some portion of permanent shock, which might be interpreted as a break in some sense.

## 5. Occasional breaks in the stock market and long memory

In this section, we analyze the long-term properties of absolute stock returns from January 4, 1928 to October 30, 2002, with 19,868 daily observations. We plot absolute returns in Fig. 4. The correlogram of absolute stock returns in Fig. 4(2) declines steadily but not exponentially. It starts with $\rho_T(1) \approx 0.4$, say, and then declines only slowly from this value. We observe another stylized fact of absolute returns, that is, the correlogram is low but remains positive for many lags in Fig. 4(2). Granger and Ding (1996) suggest a fractionally integrated model amongst the models known to generate series having such properties. By the GPH method, $\hat{d}$ is 0.450 with $t$-value 7.997. However, it can be seen from Fig. 4(1) that large absolute returns are more likely to be followed by large absolute returns than small returns. As a preliminary analysis, we plot the 240 days moving average of absolute returns in Fig. 4(3). In the beginning of the series, for example, the levels of absolute returns are significantly higher than the rest of the period. There would exist several competing models for the explanation of the long memory properties, and also time-varying $d$ in the absolute stock returns. In this section, we examine two alternative models, an occasional break model and the $I(d)$ model as a possible generating mechanism of long memory property.

The long memory in absolute stock return due to structural breaks was already investigated by Lobato and Savin (1998). They examine if the observed evidence of long memory is due to the structural break of 1973 during their sample (1962–1994). Instead of Lobato and Savin’s approach that uses a pre-determined break, we estimate unknown structural breaks in the stock market and check whether or not these breaks contain a long memory component. In our analysis, the structural breaks are estimated by Bai’s (1997) method and the absolute stock returns are decomposed into break components and residuals.

For 12 subperiods of S&P 500 daily absolute returns, Table 5 provides the estimated $d$ with $t$-statistics and a number of breaks. Fig. 5 presents plots of absolute stock returns for each subperiod. The solid line shows the sudden changes detected by plotting the mean of absolute returns, where the mean was calculated for the observations between two consecutive break points. The second column of Table 5 presents the estimated values of $d$ by GPH in the absolute returns. The values of $\hat{d}$’s range from 0.352 in period 1928–1934 to 0.154 in period 1954–1960 and up to 0.715 in period 1973–1979. All of the subperiods have strong evidence of long memory in the absolute stock return. The estimated numbers of

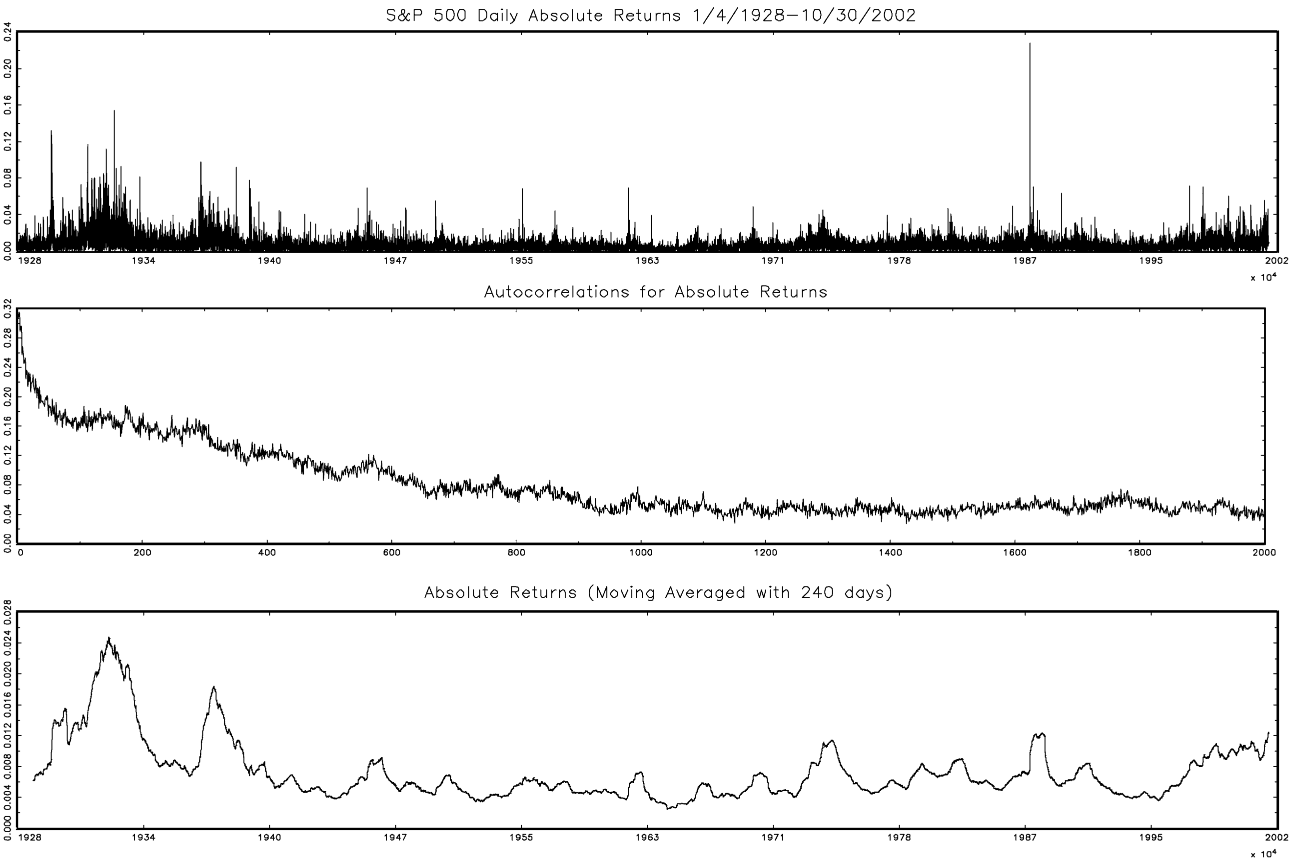


Fig. 4. (1) S&P 500 daily absolute returns 1/4/1928–10/30/2002. (2) Autocorrelations for absolute returns. (3) Absolute returns (moving averaged with 240 days).

Table 5
Memory parameter and number of breaks in absolute stock returns

| | Period | $\hat{d}$ | $\hat{R}$ | $\hat{d}$ of $(y_t - \hat{m}_t)$ |
|---|---|---|---|---|
| 1 | 1928–1934 | 0.352 (2.72) | 8 | − 0.012 ( − 0.08) |
| 2 | 1934–1940 | 0.405 (3.10) | 7 | − 0.161 ( − 1.44) |
| 3 | 1941–1947 | 0.438 (4.51) | 5 | 0.091 (0.79) |
| 4 | 1947–1953 | 0.347 (2.31) | 6 | 0.100 (1.05) |
| 5 | 1954–1960 | 0.154 (1.49) | 5 | − 0.177 ( − 1.55) |
| 6 | 1960–1966 | 0.451 (4.11) | 7 | − 0.146 ( − 1.54) |
| 7 | 1967–1973 | 0.517 (5.96) | 8 | − 0.123 ( − 1.01) |
| 8 | 1973–1979 | 0.715 (6.80) | 13 | − 0.236 ( − 2.02) |
| 9 | 1980–1986 | 0.418 (3.80) | 5 | − 0.005 ( − 0.05) |
| 10 | 1986–1991 | 0.352 (5.00) | 4 | 0.053 (0.69) |
| 11 | 1991–1998 | 0.353 (2.80) | 6 | − 0.152 ( − 1.86) |
| 12 | 1998–2002 | 0.379 (3.05) | 4 | 0.136 (0.90) |

Values in the table are calculated from 10 subsamples with sample size 1705. The 12th subsample contains only 1113 observation. The memory parameter $d$'s are estimated by the GPH method. $t$-statistics for $d$ are given in parentheses. Occasional breaks are estimated by Bai's method: To obtain consistent estimates for the error variance, we assume the number of breaks is 20 in the first step, where the size of test is chosen to be 0.05 with corresponding critical values 10.61 with $\epsilon = 0.01$.

breaks in the level of absolute returns by Bai's method (in the third column of Table 5) have positive relation with $\hat{d}$. For example, the period 1968–1973 has 13 breaks and has the highest value of $\hat{d} = 0.715$ amongst all subperiods. But they do not show an exact relation since $d$ may also be affected by the magnitude of breaks as shown in Section 3. The last column shows the estimated $d$ after filtering out structural breaks that correspond to the points of level shifts in the absolute returns. Although all of the subperiods have strong evidence of long memory in the absolute returns, residuals $\{y_t - \hat{m}_t\}$ do not have any long memory in Table 5. However, there is some possibility of overdifference as pointed out in Section 4 since many of $\hat{d}$ are negative. An alternative way to explain the possibility of overdifference is a nonlinearity, such as a smooth transition, a nonlinear trend, etc. The endogenous estimation method of multiple break points is prone to overfitting and may overestimate the number of breaks present in a given sample.

To our knowledge, currently no formal test is available for detecting multiple structural changes in the $I(d)$ process with unknown number of breaks. Alternatively, we evaluate two models by checking how well the model can explain the data. We estimate both of the models for the first 1000 observations of each subperiod. We report some of the diagnostic statistics based on residuals from estimated models. $I(d)$ models are estimated by the GPH method. We estimate breaks by Bai's (1997) repartition method in the occasional break model. We set 20 observations as the minimum length between breaks, which is equivalent to $\epsilon = 0.02$. After filtering out those components of each model from the series, we estimate the AR structure in the filtered series. The numbers of order in AR are selected by minimizing BIC. In Table 6, we report the Box-Pierce portmanteau test statistics based on the first 10 lags of the residuals, BIC and the ratio of $s_b^2/s_d^2$; $s_b$ is the residual standard deviation of the occasional break model and $s_d$ is the corresponding statistic for the $I(d)$ model. It is seen that the error variance of occasional break model is a little smaller than that of the $I(d)$ model. In Table 6, the ratios in the 8th column are very close to one, except the last period from 1986 to 1991.

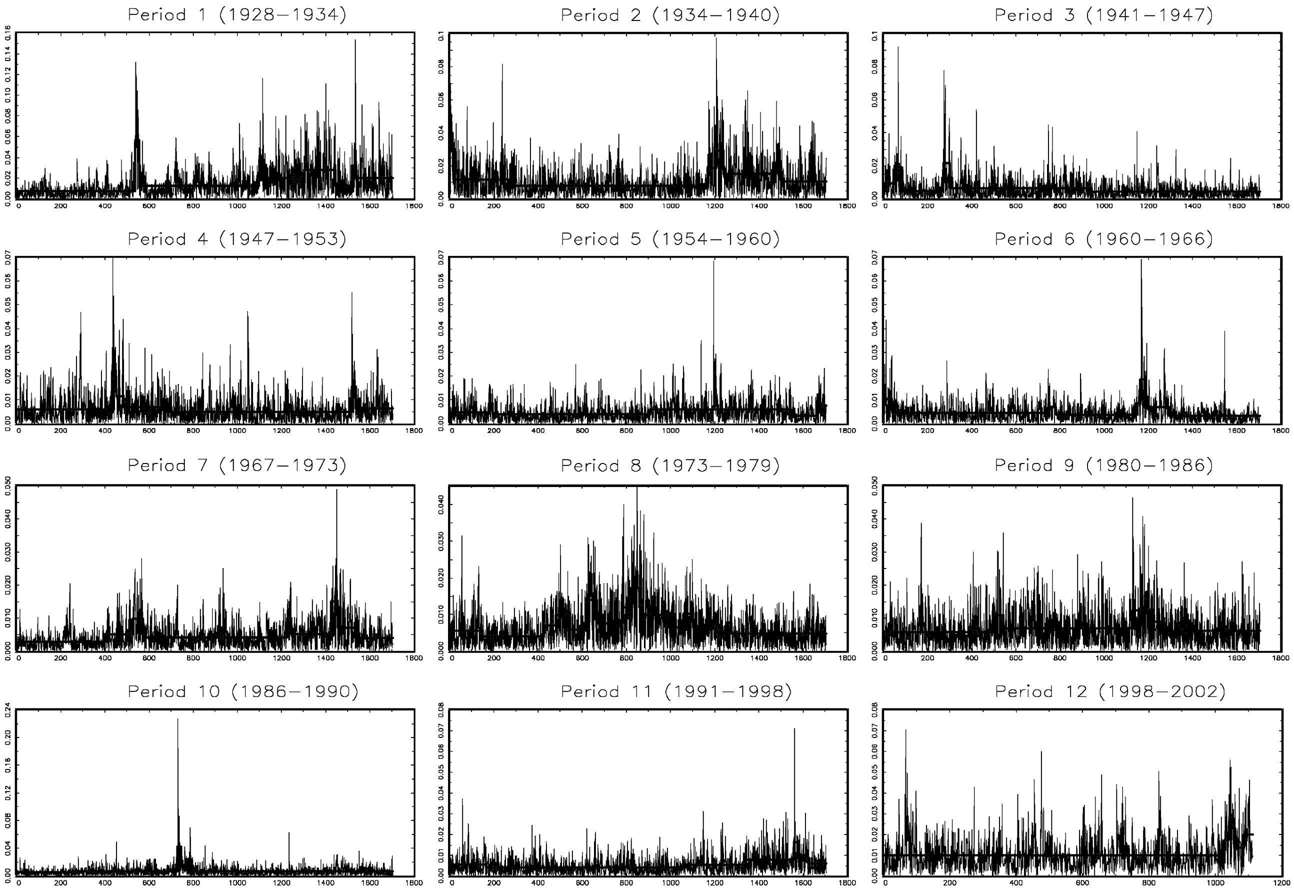


Fig. 5. (1) Period 1 (1928–1934). (2) Period 2 (1934–1940). (3) Period 3 (1940–1947). (4) Period 4 (1947–1953). (5) Period 5 (1954–1960). (6) Period 6 (1960–1966). (7) Period 7 (1967–1973). (8) Period 8 (1973–1979). (9) Period 9 (1980–1986). (10) Period 10 (1986–1990). (11) Period 11 (1991–1998). (12) Period 12 (1998–2002).

Table 6
In-sample comparison of occasional break model vs. $I(d)$ model in S&P 500 absolute stock returns

| Period | Break model | | | $I(d)$ model | | | Ratio of $s_b^2/s_d^2$ |
|---|---|---|---|---|---|---|---|
| | $Q(10)$ | BIC | LM | $Q(10)$ | BIC | SupF | |
| 1 | 4.25 | −9.23 | 0.03 | 27.77 | −9.19 | 1.86 | 0.90 |
| 2 | 3.31 | −9.57 | 0.83 | 0.66 | −9.58 | 6.20 | 1.01 |
| 3 | 0.34 | −9.62 | 2.66 | 0.39 | −9.62 | 9.96 | 0.98 |
| 4 | 21.15 | −10.05 | 0.15 | 6.09 | −10.02 | 2.61 | 0.94 |
| 5 | 7.00 | −11.19 | 0.01 | 8.54 | −11.18 | 4.72 | 0.97 |
| 6 | 0.89 | −11.04 | 0.06 | 1.18 | −11.04 | 11.88 | 0.98 |
| 7 | 7.99 | −11.20 | 0.00 | 12.60 | −11.19 | 1.50 | 0.95 |
| 8 | 13.21 | −10.20 | 0.04 | 0.15 | −10.13 | 3.73 | 0.94 |
| 9 | 28.64 | −10.36 | 3.49 | 5.17 | −10.35 | 2.39 | 1.01 |
| 10 | 3.49 | −9.24 | 4.67 | 27.73 | −9.19 | 2.42 | 0.87 |
| 11 | 9.23 | −10.97 | 1.33 | 14.71 | −10.94 | 4.74 | 0.98 |
| 12 | 12.29 | −9.48 | 1.28 | 6.83 | −9.48 | 4.01 | 1.00 |

Twelve subsamples with sample size 1705. The 12th subsample contains only 1113 observation. After estimating each model using the first 1000 in-sample observations, we calculate residuals from each model. The minimum number of observations for each regime is set at 20, i.e., $\epsilon = 0.02$. SupF denotes the sup-type break test for one more breaks. Five percent critical values for LM and SupF are 5.02 and 10.26, respectively. $s_b^2/s_d^2$: $s_b$ is the residual standard deviation of breaks model and $s_d$ is the corresponding statistic for the $I(d)$ model. $Q(10)$ are the Box–Pierce portmanteau test statistics based on the first 10 lags of the residuals.

Also, the BICs are very similar. These two models have virtually the same explanatory power for the in-sample period.

Either model by itself may not capture all of the persistence in the absolute returns, i.e., there may be residual $I(d)$ effects when a model is fitted that includes breaks only, or there may still be sudden changes in the absolute returns even after fitting an $I(d)$ model. We also examine this possibility in Table 6. The third and sixth columns contain LM statistics from the residuals of the break model and the SupF statistics (refer to Bai, 1997 for details) in the residuals of the $I(d)$ model. These statistics may provide some insights whether $d>0$ in the residuals of the break model, and whether there is a break in the residuals of the $I(d)$ model. Even though the distributions of these statistics are unknown for the estimated residual series, we may use the 5% critical value 5.02 (from $\chi_1^2$ distribution) for the LM test and 10.26 (which is 5% critical value for $\epsilon = 0.02$, see footnote 1) for the SupF for some guideline. We found that there is little evidence of $I(d)$ in the residuals after fitting the occasional breaks model, but there is only marginal evidence for an additional break in the residuals of the $I(d)$ model for period 6. In most of the subsamples, both models can explain equally well and nothing seems to be left to be explained by the other model. Therefore, a more complete analysis would exploit the possibility that the occasional break model and the $I(d)$ model can be summarized into one single model, which would be an interesting future topic.

Another way of evaluating two models is post-sample forecasting. Forecasts are obtained by estimating rolling models. Each model is estimated initially over the first 1000 observations of each subsamples, and forecasts are made for the next day, say day $T+1$, using the in-sample parameter estimates. The models are then rolled forward 1 day, deleting the first observation and adding on the observation at time $T+1$, re-estimated and a forecast is made for time $T+2$. This rolling method is repeated until the end of the out-of-sample forecast period. The one-step-ahead forecasts provide predictions for 705 out-of-sample period of each subsample. The forecasts of the occasional break model are based on the estimated mean of the last regime. The results of our forecasts are presented in Table 7. The fourth and fifth columns contain results of testing the hypothesis that there is no difference in the prediction accuracy between the two models. We applied the Fisher $z$ type test, which is based on the correlation coefficient of the sum and the differences of the forecasts errors. The last column presents the test statistics suggested by Diebold and Mariano (1995). As similar to the results of in-sample analysis, the null hypothesis was not rejected in any of the subsamples except Period 6 by $z$ test. Again, the test result of Period 6 by Diebold and Mariano's (1995) statistics shows no difference. Overall, the $I(d)$ model has marginally better forecastability than the occasional break model, although the occasional break models have marginally better explanatory power for in-sample data. To determine which model provides more information about future value, we run an encompassing regression:

$$|r_{t+1}| = \alpha + \beta_d vol^d_{t+1,t} + \beta_b \mathrm{vol}^b_{t+1,t} + \varepsilon_t$$

where we denote long-memory and break model prediction of future volatility by $\mathrm{vol}^d_{t+1,t}$ and $\mathrm{vol}^b_{t+1,t}$. If the forecasts by break model have no additional information about future volatility over that in the predicted volatility by $I(d)$ model, we would expect the coefficient $\beta_b$ to be close to zero and statistically insignificant. These results are reported in the last two columns of Table 7. There are evidence that $\beta_b$ remains significant for many cases. In some cases, e.g., for periods 2 and 9, $\beta_d$ is not significant, but $\beta_b$ is highly significant. Our results imply that a break model provides additional insight into understanding and modeling the behavior of the volatility process.

Table 7
Out-of-sample forecastability in absolute stock returns

| Period | 1-step | MSFE | Test statistics | | Encompassing test | |
|---|---|---|---|---|---|---|
| | $I(d)$ | Break | $z$ | DM | $\beta_d$ | $\beta_b$ |
| 1 | 0.1942 | 0.1963 | − 1.58 | − 1.52 | 0.62 (0.21) | − 0.10 (0.21) |
| 2 | 0.1209 | 0.1213 | − 0.39 | − 0.26 | 0.18 (0.24) | 0.55 (0.19) |
| 3 | 0.0446 | 0.0448 | − 1.22 | − 0.87 | 1.03 (0.30) | − 0.89 (0.37) |
| 4 | 0.0576 | 0.0582 | − 1.15 | − 0.92 | 0.45 (0.17) | 0.25 (0.16) |
| 5 | 0.0543 | 0.0551 | − 1.42 | − 1.26 | 0.56 (0.26) | 0.28 (0.15) |
| 6 | 0.0512 | 0.0522 | − 2.29 | − 1.15 | 0.58 (0.22) | 0.25 (0.19) |
| 7 | 0.0459 | 0.0462 | − 0.75 | − 0.57 | 0.58 (0.24) | 0.35 (0.23) |
| 8 | 0.0427 | 0.0427 | 0.005 | 0.005 | 0.36 (0.14) | 0.31 (0.16) |
| 9 | 0.0609 | 0.0609 | 0.000 | 0.000 | 0.14 (0.22) | 0.61 (0.18) |
| 10 | 0.0620 | 0.0629 | − 1.42 | − 1.18 | 0.40 (0.16) | 0.06 (0.15) |
| 11 | 0.0601 | 0.0613 | − 2.16 | − 2.01 | 0.65 (0.21) | 0.09 (0.16) |
| 12 | 0.1211 | 0.1223 | − 0.68 | − 0.61 | 1.22 (1.24) | − 0.45 (0.94) |

Twelve subsamples with sample size 1705. Values in the table are calculated from 705 one-step-ahead forecasts. The 12th subsample contains only 1113 observation and the number of out-of sample is 113. Each of the forecasts was made using estimated parameters on the 1000 previous observations. $z$ denotes the Fisher's $z$ type test statistics, and DM are the nonparametric predictive accuracy test statistics by Diebold and Mariano (1995). These two statistics are distributed standard normal. The values in MSFE are multiplied by 10. The values in the column $\beta_d$ and $\beta_b$ refer to coefficients for volatility forecasts of long memory model and break model, respectively. Standard errors of the coefficients are in parentheses.

## 6. Conclusions

In this paper, we have shown that a time series with breaks can mimic some of the properties of $I(d)$ processes where $d$ can be a fraction, its value depending on the number of breaks for a particular sample size.[4] This process has long-memory rather than short-memory, if we just consider linear properties of the data, such as autocorrelation. From the theory and simulation results, it is also shown that it is not easy to distinguish the long memory property between the break model and the one from the $I(d)$ model. In empirical analysis, we found that both of the models, the occasional break model and the $I(d)$ model, can equally well explain the absolute stock returns series. Absolute returns series may show the long memory property because of the presence of neglected breaks. Although an occasional breaks model yields a little less accurate forecasts, it can hardly be rejected on the basis of its forecasting performance relative to that of the $I(d)$ model. It will be of interest in future work to choose a plausible alternative model with breaks against $I(d)$ based on statistical criteria and economic interpretability. This has relevance for the forecastability of absolute returns, which are potentially useful for value at risk estimates, especially if the timing and the size of breaks can be shown to be forecastable. This is potentially possible if the breaks are endogenous and need to be explored further.

## Acknowledgements

This research was partly financed by NSF grant SBR-9708615. The comments from Ingolf Dittmann, Casper de Vries, Pieter van Foreest and the participants of Econometric Society Meeting, January 2001, were greatly appreciated. We also thank the two anonymous referees and the editor for many helpful comments and suggestions that result in a substantial improvement of this paper. The responsibility for any errors remains ours.

## References

Bai, J., 1997. Estimating multiple breaks one at a time. Econometric Theory 13, 315–352.

Bai, J., 1998. A note on spurious breaks. Econometric Theory 14, 663–669.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.

Baillie, R.T., 1996. Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5–59.

Chen, C., Tiao, G.C., 1990. Random level-shift time series models, ARIMA approximations, and level-shift detection. Journal of Business and Economic Statistics 8, 83–97.

Diebold, F.X., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105, 131–159.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Statistics 13, 253–263.

[4] On the completion of the first version of this paper in 1999, we found the contemporaneous and independent works by Mikosch and Stărică (1999) and Diebold and Inoue (2001) which obtain some similar results to ours, but with somewhat different implications. In many ways the two sets of results complement rather than compete with each other.

Engle, R.F., Smith, A.D., 1999. Stochastic permanent breaks. Review of Economics and Statistics 81, 553–574.
Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 15–39.
Granger, C.W.J., 1998. Real and spurious long-memory properties of stock market data: comment. Journal of Business and Economic Statistics 16, 268–269.
Granger, C.W.J., Ding, Z., 1995a. Some properties of absolute returns. An alternative measure of risk. Annales d'Economie et de Statistique 40, 67–91.
Granger, C.W.J., Ding, Z., 1995b. Stylized Facts on the Temporal and Distributional Properties of Daily Data from Speculative Markets. UCSD Working Paper.
Granger, C.W.J., Ding, Z., 1996. Varieties of long memory models. Journal of Econometrics 73, 61–78.
Granger, C.W.J., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis 1, 15–39.
Granger, C.W.J., Marmol, F., 1997. The Correlogram of a Long Memory Process Plus a Simple Noise. UCSD Working Paper.
Granger, C.W.J., Teräsvirta, T., 1999. A simple nonlinear time series model with misleading linear properties. Economics Letters 62, 161–165.
Kuan, C.-M., Hsu, C.-C., 1998. Change-point estimation of fractionally integrated process. Journal of Time Series Analysis 19, 693–708.
Lavielle, M., Moulines, E., 2000. Least-squares estimation of an unknown number of shifts in a time series. Journal of Time Series Analysis 21, 33–59.
Lobato, I.N., Robinson, P.M., 1998. A nonparametric test for I(0). Review of Economic Studies 65, 475–495.
Lobato, I.N., Savin, N.E., 1998. Real and spurious long-memory properties of stock-market data. Journal of Business and Economic Statistics 16, 261–268.
Mikosch, T., Stărică, C., 1999. Change of structure in financial time series, long range dependence and the GARCH Model, Working paper. The Wharton School, Chalmers University of Technology.
Nunes, L.C., Kuan, C.-M., Newbold, P., 1995. Spurious Break. Econometric Theory 11, 736–749.
Nunes, L.C., Newbold, P., Kuan, C.-M., 1996. Spurious number of breaks. Econometrics Letters 50, 175–178.
Robinson, P.M., 1995. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048–1072.
Rydén, T., Teräsvirta, T., Åsbrink, S., 1998. Stylized facts of daily returns series and hidden Markov model. Journal of Applied Econometrics 13, 217–244.
Yao, Y.-C., 1988. Estimating the number of change-points via Schwarz' criterion. Statistics & Probability Letters 6, 181–189.