STRUCTURAL CHANGE POINT TESTING WITH APPLICATION TO STOCK RETURNS

JONATHAN BOWN

Supervisory Committee:

Lajos Horváth (Committee Chair)

Tom Alberts

Ron Reeder

A project submitted to the faculty of

The University of Utah in partial fulfillment of the requirements for the degree of

Master of Statistics

Department of Mathematics

Abstract. We propose semi–parametric tests to detect a change point in the structure or mean of sequences of independent random variables and time dependent processes. The asymptotic distributions of the proposed statistics are derived under moderate assumptions. We discuss the applicability of our method to test for parameter changes to the most often used distributions and models, including autoregressive moving average (ARMA) and autoregressive conditional heteroskedastic (ARCH). Our simulations show that our tests have good size and power properties for independent and dependent sequences of random variables. We apply the semi–parametric tests in the attempt to date the occurrence of major declines during the Great Recession and several other major financial events. We discuss limitations with these approaches as well as possibilities for further research.

Contents

1. Introduction

In the real world, data does not often maintain the same statistical properties over time or across samples. One purpose of investigating changes in the statistical properties is to find when the change actually occurred. This can be a challenge when seeking a change point in real-time or possibly in the future because the amount of data available becomes more limited. This leads to the question of how long must one wait or how much data to collect to find an accurate change point and perhaps exit or enter a financial asset based on this information. This question can be examined with a wide variety of tools and methods with everything from prediction to parameter estimation but can also be burdened with assumptions. The test we propose is an effort to simplify the assumptions needed to perform a change point test in addition to cutting down on the information that is required to find an adequate change point. The semi-parametric test is also flexible to apply weak or strong parametric assumptions about future examples.

This paper is structured as follows. Section 2 introduces the semi–parametric test and its properties. Section 3 shows our choice of boundary function used as part of the test to achieve convergence and satisfy desired properties of the test. Section 4 provides details of the simulations that were performed to derive the critical values. Section 5 provides examples of the parametric approach for which the assumptions of our theoretical framework are satisfied. In Section 6, we assess the finite sample performances of the proposed tests on dependent sequences and provide an estimator used for the parameter in the test statistic in parametric and non-parametric settings. Section 8 provides an empirical application in the context of stock returns for various stocks that have historically exhibited a severe change in asset price or asset return. The R code utilized for this project is located in a GitHub repository referenced by [4].

2. Proposed Test for Change Points

We write the measurement error model as $X_i = \mu_i + e_i$ for $i \geq 1$, where X_i is the i^{th} observation and the errors e_i are stationary with $E[e_i] = 0$ for $i \geq 1$. We have a training sample of size M with a stable but unknown mean $\mu_i = \mu$ for $1 \leq i \leq M$. We are interested in testing if a change at an unknown k^* occurs when we continue sampling data after the initial sample. We have the following model for the means

$$\mu_i = \begin{cases} \mu, & M+1 \leq i \leq M+k^*, \\ \mu + \Delta, & i \geq M + k^* + 1, \end{cases}$$

where $\Delta \neq 0$ and unkown. The total number of observations we collect after the training sample is T. The null hypothesis for this setting is given by

$$H_0 : k^* > T \;\; (\mu_1 = \mu_2 = ... = \mu_M = \mu_{M+1} = ... = \mu_{M+T}),$$

i.e. there is no change in the mean during the observation period. The alternative hypothesis is given by

$$H_A : k^* < T \quad (\mu_1 = ... = \mu_M = .. = \mu_{M+k^*} \neq \mu_{M+k^*+1} = ... = \mu_{M+T}),$$

i.e. there was a change in the mean and it occurred at data point k^*. We are interested in a sequential procedure to get H_0 against H_A. Since our procedure will terminate at time T after the training period, our method is a long end procedure. This model was introduced by Aue and Horváth (2012) with a survey on sequential methods, and their applications. Our procedure is based on estimating and comparing the mean of the elements of the test (current) sample to the mean of the training (historical) sample. After we collected additional $k+h$ observations, we compare the mean of $X_{M+k}, ..., X_{M+k+h}$ to the sample mean of the training sample. Usually, h is called the "rolling window" in econometrics. Let

$$Z_k = \left| \overline{X}_M - \overline{X}_{k,h} \right|,$$

where

$$\overline{X}_{k,h} = \frac{1}{h} \sum_{i=M+k}^{M+k+h} X_i, \tag{2.1}$$

and the stopping time will be defined as

$$\tau_M = \min \left\{ k < T - h : Z_k > g_\alpha(h,k), \; T - h \right\}. \tag{2.2}$$

We are interested in detecting when the quantity Z_k crosses a suitably chosen boundary function $g_\alpha(h,k)$. Note that (2.1) is the sample mean of the observations, $X_{M+k}, ..., X_{M+k+h}$. This means that we collected already $M+k+h$ observations after the training period and only the last h are compared to $\overline{X}_M$. This method should be better than comparing the sample mean of $X_{M+1}, ..., X_{M+k+h}$ to the mean of the training sample. If the change comes late, then several additional observations have been already collected with the mean of the training sample. Hence a large number of observations after the time of change must be collected to overcome their effect in the sample mean of the observations collected after the training period. We require that

$$\lim_{M\to\infty} \mathrm{P}(\tau_M < T - h) = \alpha, \tag{2.3}$$

i.e. the probability of false alarm is α. This means that the probability of the type I error is α for large M. Under the alternative hypothesis there is exactly one change. We also require that we stop collecting further observations if there is a change in the data and we wish to stop as soon as possible. We wish to have under the alternative that

$$\lim_{M\to\infty} \mathrm{P}(\tau_M < T - h) = 1,$$

i.e. we always detect the change with very large probability. We note that under the null hypothesis and the assumed stationarity of the errors $\{e_i, i \geq 1\}$, the variance of (2.1) does not depend on k nor on M. We choose c such that (2.3) holds under H_0. Throughout this project we use the notation

$$\xi_n = o(a_n) \ \text{ a.s. if } \lim_{n\to\infty} \frac{|\xi_n|}{a_n} = 0 \ \text{ a.s.,}$$
$$\xi_n = O(a_n) \ \text{ a.s. if } \lim_{n\to\infty} \frac{|\xi_n|}{a_n} < \infty \ \text{ a.s.}$$

Similarly,

$$\xi_n = o(a_n) \ \text{ in probability } \ (\xi_n = o_{\mathrm{P}}(a_n))$$

means that

$$\limsup_{n\to\infty} \mathrm{P}(|\xi_n|/a_n \geq \varepsilon) = 0 \ \text{ for all } \varepsilon > 0$$

and

$$\xi_n = O(a_n) \ \text{ in probability } \ (\xi_n = O_{\mathrm{P}}(a_n))$$

if

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathrm{P}(\xi_n/a_n \geq K) = 0.$$

We write

$$\varepsilon \sim N(0,1)$$

to say that the random variable ε has a normal distribution with mean zero and variance one. We use the following assumption to arrive at our results.

Assumption 2.1. *There are partial sums of stationary random variables $\{e_i, i \geq 1\}$ and Wiener processes $\{W_1(u), u \geq 0\}$ and $\{W_2(u), u \geq 0\}$ such that*

$$\sum_{i=1}^{M} e_i = \sigma W_1(M) + o(M^{\epsilon}) \ \text{ a.s. for some } \ 0 < \epsilon < 1/2 \ \text{ as } \ M \to \infty, \tag{2.4}$$

(2.5) $$\sum_{i=M+1}^{k+M} e_i = \sigma W_2(k) + o(k^{\epsilon}) \ \ \textit{a.s. for some} \ \ 0 < \epsilon < 1/2, \ \ \textit{as } k \to \infty,$$

and

(2.6) $$\{W_1(s), 0 \le s \le M\} \ \textit{and} \ \{W_2(t), 0 \le t < \infty\} \ \textit{are independent.}$$

These assumptions require that $\mathrm{E}|e_i|^{1/\epsilon} < \infty$. We note that in our calculations and in the computation of the critical values with the boundary function $g_\alpha(h,k)$ it is easier to work with a continuing formulation of (2.5). According to Csörgő and Révész (1981) for any Wiener process $\{W(u), u \ge 0\}$

(2.7) $$\sup_{k \le t \le k+1} \frac{|W(k) - W(t)|}{\sqrt{\log(k)}} < \infty \ \text{ a.s., as } k \to \infty,$$

and therefore (2.5) is equivalent with

(2.8) $$\sum_{i=M}^{t+M} e_i = \sigma W(t) + o(t^{\epsilon}) \ \text{ a.s.}$$

We assume that $\{e_i, i \ge 1\}$ is a stationary process, such as the ARMA(p,q), i.e.

(2.9) $$e_t = \delta_1 e_{t-1} + ... + \delta_p e_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} \quad -\infty < t < \infty,$$

where $\{\varepsilon_t, -\infty < t < \infty\}$ are independent and identically distributed random variables with $E[\varepsilon_0] = 0$ and $E|\varepsilon_0|^\nu < \infty$ for some $\nu > 2$. It is shown in Shumway and Stoffer (2017) that (2.9) has a stationary solution if and only if the roots of the polynomials $\delta(z), \theta(z)$ are outside of the unit circle. According to Khoshnevisan (2013), if $E|\varepsilon_0|^\nu < \infty$, then

(2.10) $$\limsup_{k \to \infty} \frac{|\varepsilon_k|}{k^{1/\nu}} = 0 \ \text{ a.s.}$$

The stationary solution can be written as

$$e_s = \sum_{l=0}^{\infty} c_l \varepsilon_{s-l},$$

where $|c_l| = o(\delta^l)$ as $l \to \infty$ with some $0 < \delta < 1$. Hence by Minkowski's inequality we have

$$\begin{aligned}(E|e_s|^\nu)^{1/\nu} &= \left(E \left| \sum_{l=0}^{\infty} c_l \varepsilon_{s-l} \right|^\nu \right)^{1/\nu} \\ &\le \sum_{l=0}^{\infty} \left(E|c_l \varepsilon_{s-l}|\right)^{1/\nu} \\ &\le \sum_{l=0}^{\infty} |c_l| \left(E|\varepsilon_0|\right)^{1/l} < \infty.\end{aligned}$$

Since $\{e_i, i \geq 1\}$ is a stationary sequence with $E|e_0|^\nu < \infty$, using the Borel-Cantelli lemmas as in Khoshnavishan (2013) we get along the lines of (2.10) that

$$\limsup_{k\to\infty} \frac{|e_k|}{k^{1/\nu}} = 0 \;\; \text{a.s.} \tag{2.11}$$

Now we take the sums of (2.9) and get

$$\sum_{t=1}^{M} e_t = \delta_1 \sum_{t=1}^{M} e_{t-1} + ... + \delta_p \sum_{t=1}^{M} e_{t-p} + \sum_{t=1}^{M} e_t + \theta_1 \sum_{t=1}^{M} \varepsilon_{t-1} + \cdots + \sum_{t=1}^{M} e_{t-q}$$

$$= (\delta_1 + ... + \delta_p) \sum_{t=1}^{M} \varepsilon_t + R_M^{(1)} + (1 + \theta_1 + ... + \theta_q) \sum_{t=1}^{t=M} \varepsilon_t + R_M^{(2)}, \tag{2.12}$$

where

$$R_M^{(1)} = \delta_1(e_0 - e_M) + \cdots + \delta_p \left[(e_{1-p} + e_{2-p} + \cdots + e_0) - (e_M + e_{M-1} + \cdots + e_M + 1)\right],$$

and

$$R_M^{(2)} = \theta_1(e_0 - e_M) + \cdots + \theta_q \left[(e_{1-p} + e_{2-q} + \cdots + e_0) - (e_M + e_{M-1} + \cdots + e_M + 1)\right].$$

It follows from (2.11) that

$$\lim_{M\to\infty} \frac{|R_M^{(1)}|}{M^{1/\nu}} = 0 \;\; \text{a.s.}$$

and (2.10) yields

$$\limsup_{M\to\infty} \frac{|R_M^{(2)}|}{M^{1/\nu}} = 0 \;\; \text{a.s.}$$

Hence

$$\sum_{t=1}^{M} e_t = \frac{1 + \theta_1 + \cdots + \theta_q}{1 - (\delta_1 + \cdots + \delta_p)} \sum_{t=1}^{M} \varepsilon_t + o(M^{1/\nu}) \quad \text{a.s., as } M \to \infty. \tag{2.13}$$

Using the Komlós-Major-Tusnády approximation (Csörgő and Révész (1981)) there is a Wiener process $\{W_1(u), u \geq 0\}$ such that

$$\sum_{t=1}^{M} \varepsilon_t - (\mathrm{Var}(\varepsilon_0))^{1/2} W_1(M) = o(M^{1/\nu}) \quad \text{a.s., as } M \to \infty.$$

For some $\nu > 2$. Hence (2.4) holds with

$$\sigma^2 = \left(\frac{1 + \theta_1 + \cdots + \theta_q}{1 - (\delta_1 + \cdots + \delta_p)}\right)^2 \mathrm{Var}(\varepsilon_0). \tag{2.14}$$

Similarly to (2.13) we can prove that

$$\sum_{t=M+1}^{M+k} e_t = \frac{1+\theta_1+\cdots+\theta_q}{1-(\delta_1+\cdots+\delta_p)} \sum_{M+1}^{M+k} \varepsilon_t + o(k^{1/\nu}).$$

Hence (2.6) holds by the independence of $\left\{\sum_{i=1}^{M+k} \varepsilon_i, k \geq 1\right\}$.

The approximation follows again from the Komlós-Major-Tusnády approximation. The next result gives the distribution of τ_M under the null hypothesis. Let

$$g_\alpha(h,k) = \frac{c_{1-\alpha}\sigma(h+k)^\beta}{h^{\beta+1/2}} \tag{2.15}$$

where $\beta > 1/2$ and $c_{1-\alpha}$ is the upper quantile or critical value associated with the size of the test.

Theorem 2.1. *If Assumption 2.1 holds, $h/M \to 0$, $h/T \to 0$ as $M, T \to \infty$, and $g_\alpha(h,k)$ is defined by (2.15), then*

$$\lim_{M\to\infty} \mathrm{P}\left(\tau_M < T-h\right) = \mathrm{P}\left(\sup_{0<u<\infty} \frac{|W(u+1)-W(u)|}{(u+1)^\beta} \leq c_{1-\alpha}\right),$$

where $\{W(u), u \geq 0\}$ is a Wiener process.

If H_A holds and Assumption 2.1 is satisfied, then

$$\frac{Z_k}{|\Delta|} \xrightarrow{\mathrm{P}} 1,$$

where Δ denotes the size of the change. So if

$$\limsup_{h\to\infty} \frac{k^*}{h} < \infty, \tag{2.16}$$

then

$$\frac{Z_k}{g_\alpha(h,k^*)} \xrightarrow{\mathrm{P}} \infty, \tag{2.17}$$

assuming that

$$h^{1/2}|\Delta| \to \infty. \tag{2.18}$$

Under (2.16) we have that

$$\liminf_{h\to\infty} \frac{h^{-1/2}}{g_\alpha(h,k^*)} > 0$$

and therefore (2.18) implies (2.17). If

$$\lim_{h\to\infty} \frac{k^*}{h} = \infty$$

then

$$\frac{h^*}{k^*+h} = \infty$$

so if

$$|\Delta|\, h^{1/2} \left(\frac{h}{k^*}\right)^{\beta} \to \infty \tag{2.19}$$

then (2.17) holds again. Equation 2.17 yields under H_A that

$$\lim_{M\to\infty} \mathrm{P}(\tau_M < T-h) = 1, \tag{2.20}$$

i.e. we detect the change with probability converging to 1. To achieve this we need to collect only k^*+h additional observations. This indicates that the size of the change can be small and our procedure will detect it with high probability. If β is small in (2.19), then the detectable change can also be small. The sequential procedure can be compared to the fixed sample size statistic

$$\Theta_{M,T} = \max_{1\le k\le T-h} \left| \overline{X}_M - \frac{1}{h}\sum_{i=k}^{k+h} X_i \right| \Bigg/ g_\alpha(h,k). \tag{2.21}$$

Theorem 2.2. *Let $X_i, 1 \le i \le M$, be i.i.d. random variables, also fix an $h > 0$ such that $h/M \to 0$ and $h/T \to 0$ as $M, T \to \infty$, and let $g_\alpha(h,k)$ be defined properly. Then*

$$\Theta_{M,T} \xrightarrow{\mathcal{D}} \max_{0<u<\infty} \frac{|W(u+1)-W(u)|}{(u+1)^\beta}$$

for $\beta > 1/2$ and $W(t)$ is a Wiener process, or standard Brownian motion.

Note here that $\xrightarrow{\mathcal{D}}$ refers to weak convergence or convergence in distribution. The test statistic $\Theta_{M,T}$ requires the collection of additional T observations after the training period. The sequential procedure only has T observations under the H_0 while only k^*+h under the alternative. This gives a sense that the sequential procedure is more efficient and less complex under the alternative. Under H_0 these procedures are equivalent.

Lemma 2.1. *We assume that the assumptions of Theorem 2.1 are satisfied. Then,*

$$\mathrm{P}(\tau_M < T-h) \xrightarrow{\mathcal{D}} \mathrm{P}\left(\max_{0<u<\infty} \frac{|W(u+1)-W(u)|}{(u+1)^\beta} \ge c_{1-\alpha}\right) = \alpha,$$

where $c_{1-\alpha}$ is the critical value chosen for the test of size α.

3. Brownian Motion and Boundary Functions

In this section, the finite sample performance of the monitoring procedure is evaluated. We make use of standard Brownian motion to achieve the simulation results of the critical values as well as the size of the test for $\Theta_{M,T}$. We make heavy use of standard Brownian motion. We breifly introduce the basic definition of a Wiener process or Brownian motion according to Wiener (1923) are as follows, Brownian motion $\{W(t)\}_{t\ge 0}$ is a random function of t (= "time") such that:

(1) $\mathrm{P}(W_0 = 0) = 1$.

(2) W has independent increments: for every $t > 0$, the future increments $W(t + u) - W(t),\ u \geq 0$ are independent of the past values $\{W(s),\ s < t\}$.

(3) W has Gaussian increments: $W(t + u) - W(t)$ is normally distributed with zero mean and variance u.

(4) W has continuous paths: with probability $1, W(t)$ is continuous in t.

The Wiener process is sometimes called a standard Brownian motion. By definition $\mathrm{E}[W(t)] = 0$ and $\mathrm{Var}(W(t)) = \mathrm{E}[W(t)^2] - \mathrm{E}[W(t)]^2 = \mathrm{E}[W(t)^2] - 0 = \mathrm{E}[W(t)^2] = t$ (DasGupta, 2008). In order to determine the underlying distribution of the limit of the statistic $\Theta_{M,T}$ we need the following theorems about the increment limits of standard Brownian motion. We have the following from Csörgő and Révész (1981),

Theorem 3.1. *If $W(t)$ is a Wiener process then*

$$\limsup_{t\to\infty} \frac{|W(t)|}{\sqrt{2t\log(\log(t))}} = 1 \ \ a.s.$$

Theorem 3.2. *If $W(t)$ is a Wiener process then*

$$\limsup_{t\to\infty} \frac{|W(t)|}{\sqrt{2t\log(1/t)}} = 1 \ \ a.s.$$

Theorem 3.3. *If $W(t)$ is a Wiener process and $\Gamma(y) = W(y+1) - W(y)$ then*

$$\limsup_{y\to\infty} \frac{|\Gamma(y)|}{\sqrt{2\log(y)}} = 1 \ \ a.s.$$

Theorem 3.3 in addition to Theorem 2.1 lead us to the distribution of $\Theta_{M,T}$ and then to the selection of the suitable boundary function. In particular, Theorem 3.3 gives us a baseline for convergence. If we can further bound the denominator below by $\sqrt{2\log(y)}$ and have the selected quantity be a monotone increasing function of y, then we will have convergence to zero, instead of one, in the limit of the supremum. This is achieved by using $(1+u)^{\beta}$ in the denominator of (4.1). The proof of Theorem 3.1 is simple so we provide it in Appendix B.

The limiting distribution of the test statistic depends on the choice of the constant, β. This constant is a small number chosen before the test that allows proper convergence of the test statistic. The value of β was evaluated based on the initial simulations then set as a constant for the remainder of examples and applications. The critical values for $\Theta_{M,T}$ are found initially under different values of β in order to compare the test size and determine the best value. We also evaluated how the rolling window should depend on the amount of information we are using in the test sample of size T. The intuition being that as T increases, the window h should not be as refined as when T is small which is more formally introduced in Section 4. First, we introduce the definitions of the boundary function and the resulting simulations for independent identically distributed random variables.

Definition 3.1. *Let $X_i, 0 \leq i \leq M$ be i.i.d. random variables with mean 0 and variance 1. If assumption* (2.4) *is met, fix an $h > 0$ such that $h/M \to 0$ and $h/T \to 0$, then the boundary function is given by*

$$g(h,k) = \frac{(h+k)^\beta}{h^{\beta+1/2}}$$

for $\beta > 1/2$ and $k \in \mathbb{Z}^+$.

Note that in the above definition we did not include the subscript α because the definition does not include the critical value in the function $g(\cdot)$. We want to make the distinction that this is the basic form of the boundary function in general for variance one random variables. Including the subscript α indicates the inclusion of the critical value $c_{1-\alpha}$ in the numerator of the function $g(\cdot)$ and merely a matter of notation for the analysis.

Definition 3.2. *Let $X_i, \quad 0 \leq i \leq M$ be stationary random variables with mean 0 and unknown long-run variance σ^2. If assumption* (2.4) *is met, fix an $h = h(T) > 0$ such that $h/M \to 0$ and $h/T \to 0$ as $M, T \to \infty$ then the boundary function is given by*

$$g(h,k) = \frac{\hat{\sigma}(h+k)^\beta}{h^{\beta+1/2}}$$

for $\beta > 1/2$ and $k \in \mathbb{Z}^+$.

For the i.i.d. boundary function, $\hat{\sigma} = \sqrt{S^2}$, where $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$. For dependent stationary random variables, $\hat{\sigma}$ is the non-parametric estimator of the long-run variance.

4. Simulation of Critical Values

The random variable $\Theta_{M,T}$ is a maximum on the positive half line. Of course numerically we cannot take infinitely many values so we compute

$$\Theta(u) = \sup_{0<t\leq u} \frac{|W(t+1) - W(t)|}{(t+1)^\beta}. \tag{4.1}$$

Since

$$\sup_{u \leq t < \infty} \frac{|W(t+1) - W(t)|}{(t+1)^\beta} \to 0 \text{ a.s.},$$

$\Theta(u)$ will change little after some u. Let $\varepsilon_1, \varepsilon_2, ...$ be a sequence of independent, identically distributed random variables with mean 0 and variance 1. For each $n \geq 1$ define a continuous-time stochastic process $\{W_n(t)\}_{t\geq 0}$ by

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{1\leq j \leq \lfloor nt \rfloor} \varepsilon_i, \quad 0 \leq t \leq 1. \tag{4.2}$$

This is a form of the one-dimensional random walk and is a random step function with jumps of size $\pm 1/\sqrt{n}$ at times k/n, where $k \in \mathbb{Z}^+$. Since the random variables ε_j are independent, the increments of $W_n(t)$ are independent. In addition, for large N the distribution of $W_n(t+s) - W_n(s)$ is close to the $N(0,t)$ distribution by the Central Limit Theorem. Thus as $n \to \infty$, the distribution of the random function $W_n(t)$ approaches that of standard Brownian motion. This property of the random walk was used in the simulation of the test statistic.

Using the details above, we constructed a way to simulate the necessary quantity in which $\Theta(u)$ converges to, given by Theorem 2.1. Primarily we know that we can't let u go off to infinity during the simulation. For each simulation of the critical values we needed to select an endpoint u and show that the distribution of this quantity converges as if $u \to \infty$. To do this we will select $u \in \{1,2,3,4,5,6\}$ in order to show that the critical values converge very quickly. We also need t to take on very small increments of values as if it were on a continuous interval $[0,u]$. Thus we will use a quantity dt which will be the frequency of the sequence generated on $[0,u]$ which gives an interval of length u/dt. One other technicality is that the corresponding Brownian motion is defined on the interval $[0,u+1]$. This ensured that $\Theta(u)$ was well-defined.

In the simulations of Brownian motion, we made sure that Definition 4.2 and the properties above are satisfied. We started by letting $\mathbf{E} = (\varepsilon_1, \varepsilon_1+\varepsilon_2, ..., \sum_{j=1}^{\lfloor nt \rfloor} \varepsilon_j)$. $\mathbf{E}$ is a vector that when simulated contains the cumulative sum terms of $W(t)$ for every t on the interval $[0,u]$. Let $N = \lfloor (u+1)/dt \rfloor$, this is the number of elements in $\mathbf{E}$. The last part to determine was the quantity n in (4.2). For convenience of translating these simulations in R, we had the random walk to go from index 1 to $\lfloor nt \rfloor$. We have three simple equations that must hold in order to solve for n, namely

$$t = dt \text{ s.t. } \lfloor n(dt) \rfloor = 1, \tag{4.3}$$

$$t = u \text{ s.t. } \lfloor n(u) \rfloor = \frac{u}{dt}, \tag{4.4}$$

$$t = u+1 \text{ s.t. } \lfloor n(u+1) \rfloor = \frac{u+1}{dt}. \tag{4.5}$$

When we evaluated (4.1), the numerator was the difference between the Brownian motion vector at index u/dt and $(u+1)/dt$. We needed these conditions in order to produce the correct indices as these differences were evaluated. I show here that $n = N/(u+1)$ satisfies all of these desired conditions. Recall that

$$N = \left\lfloor \frac{u+1}{dt} \right\rfloor,$$

thus

$$n = \frac{N}{u+1} = \frac{\lfloor \frac{u+1}{dt} \rfloor}{u+1},$$

and $u+1$ is an integer so we can extend the floor function as

$$\left\lfloor \frac{u+1}{dt(u+1)} \right\rfloor = \left\lfloor \frac{1}{dt} \right\rfloor.$$

In the simulations performed, we chose $dt = 0.001$ such that $1/dt = 1000$. Thus when $n = 1/dt = 1000$ we had immediately that (4.3), (4.4), and (4.5) hold. The plots and tables below begin by showing the convergence of the critical values of (4.1). These values were found by evaluating the quantiles of the $\Theta(u)$ on intervals of specified end-point. Once the quantiles (critical values) had been found to converge, we checked that the simulated size of the test is close to the chosen size under the null hypothesis. This involved picking an M and a T as in (2.21) in order to generate the training and test sets. We simulate this for i.i.d. random samples as well as dependent sequences. After the test size was evaluated, we simulated the sequential procedure to evaluate stopping times. For all these tests we compared among values of $\beta \in \{1, 2, 3, 4\}$ to find the best value that allowed the simulated size of the test to be close to the desired size of the test. Recall we claimed that the smaller the values of β are better able to detect smaller changes in the mean. The only requirement on β is that it is chosen to be greater than 1/2.

Previous to the results shown for the size of the test, there was a lot of testing of what would be the ideal rolling window, h. Initially this was chosen to be a fixed number such as $h = 2$. However, the simulations showed a lot of variability in the size of the test and in the overall sequential procedure that were not satisfactory. There was also no theoretical justification for h being a constant. If h is a constant, then there is nowhere in $\Theta_{M,T}$ that actually depends on the number of test examples, T. We have the training sample size, M, contained in the calculation of the sample mean and sample variance. We also will have more terms to take the maximum over as T grows larger. But it seems natural to include T in the boundary function in order to scale the differences by some function of how much information we have in the test sample which similar to other basic hypothesis tests. There should also be some flexibility of the rolling window for calculating the means at each increment of k depending on if there is a little or large amount of examples. This led to the idea to let $h = \lfloor T^{1/2} \rfloor$. This fits the assumptions necessary for Theorem 2.2. We have that

$$\begin{aligned} h = \lfloor T^{1/2} \rfloor &\to \infty \quad \text{as } T \to \infty, \\ \lfloor T^{1/2} \rfloor / M &\to 0 \quad \text{as } M, T \to \infty, \\ \lfloor T^{1/2} \rfloor / T &\to 0 \quad \text{as } T \to \infty. \end{aligned}$$

Throughout the simulations and applications, the boundary function is implemented as

$$g(h, k) = g(T^{1/2}, k) = \frac{\hat{\sigma}(T^{1/2} + k)^{\beta}}{T^{\beta/2 + 1/4}}, \tag{4.6}$$

and the critical values used are shown in Table 4.2. The estimator $\hat{\sigma}$ is interchangeable with σ in (4.6) and unless otherwise specified, the simulations use $\hat{\sigma}$. The above formulation of the boundary function was a key factor in obtaining consistent results for the size of the test across all the models examined. Other attempts were made at functions of perhaps M or other functions of T, but the most consistent results were a consequence of (4.6).

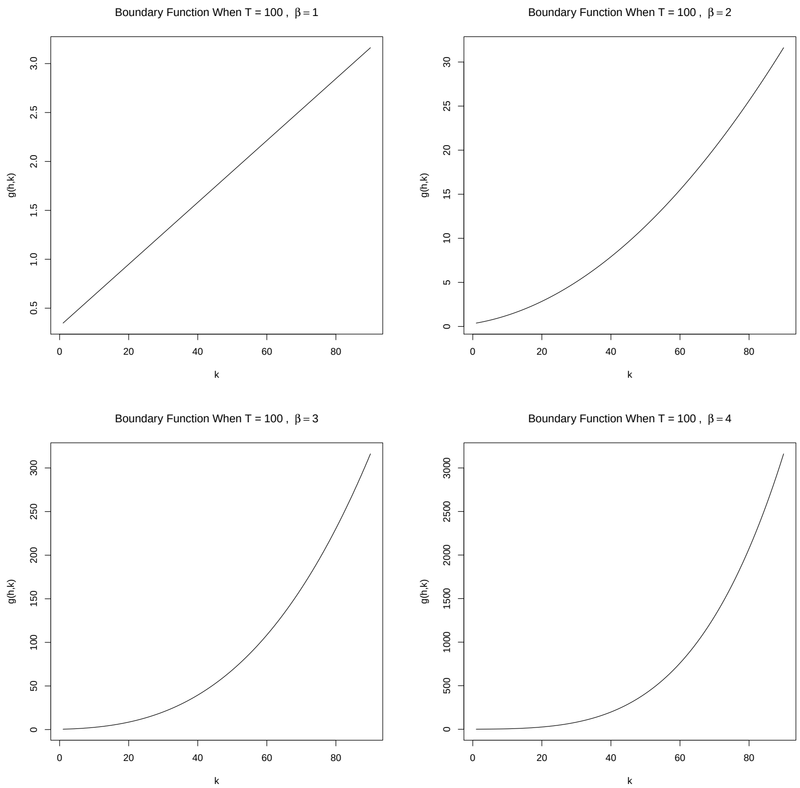


FIGURE 4.1. Boundary function for different values of β

As part of the simulations of test size, we also evaluated the consistency of this dependence on T for the value of h. To test $h = \lfloor T^{1/2} \rfloor$ we chose four scenarios for M and T. Two scenarios let M equal T for small and large values. Two scenarios give a clear imbalance between M and T. These scenarios are a way of creating a large imbalance between the test and training sample to see if this imbalance is reflected in the empirical size of the test. However, the resulting change in the size of the test is negligible as can be seen in the simulation results. The most significant factor in the size of the test was the value chosen for β. All simulations were carried out using the R software, and each example was simulated a minimum of 5,000 times. The figures that follow are plots of the density and distribution functions of $\Theta(u)$ for several values of u.

FIGURE 4.2. Simulated density and distribution function of $\Theta(1)$

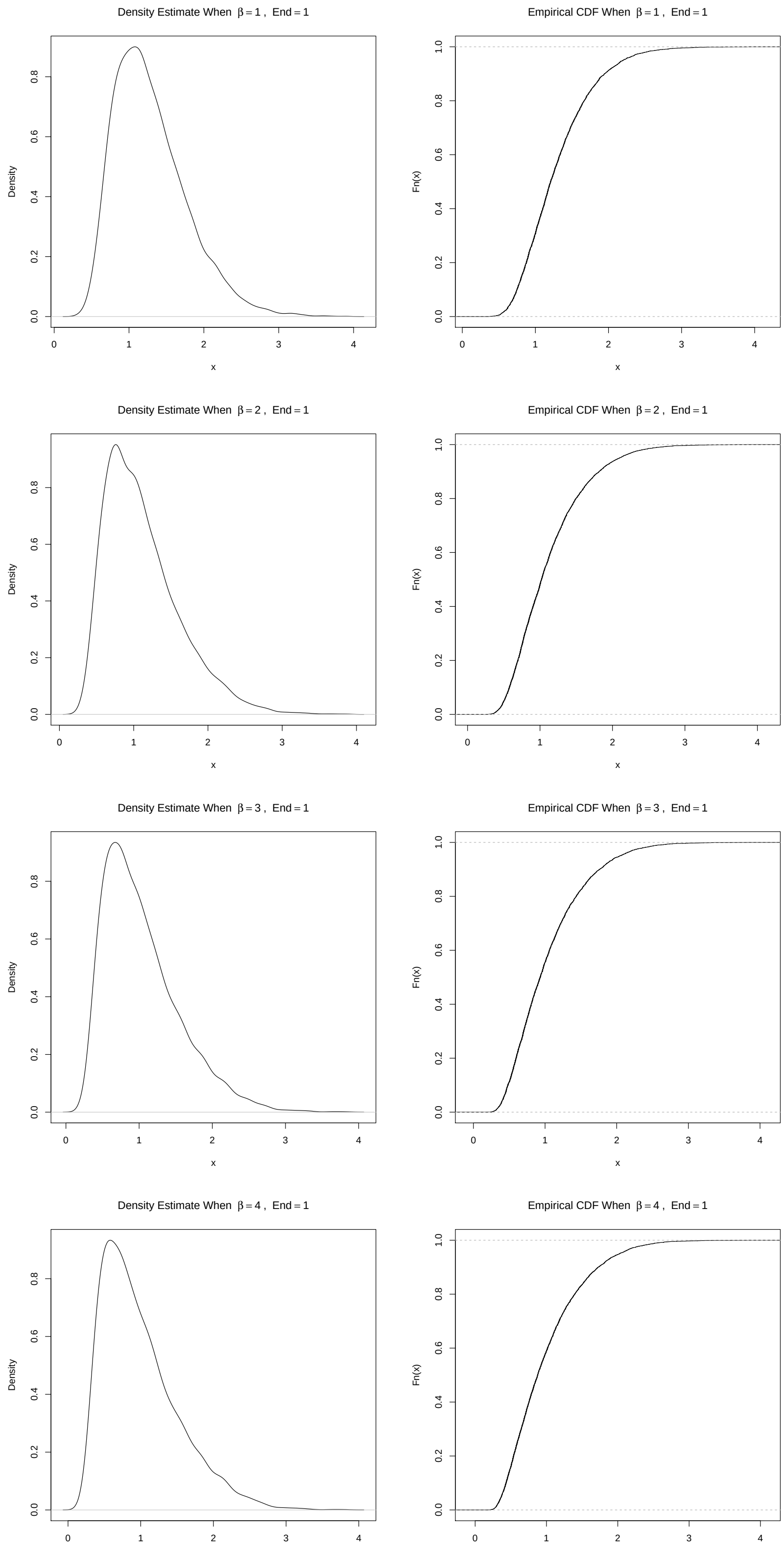


FIGURE 4.3. Simulated density and distribution function of $\Theta(2)$

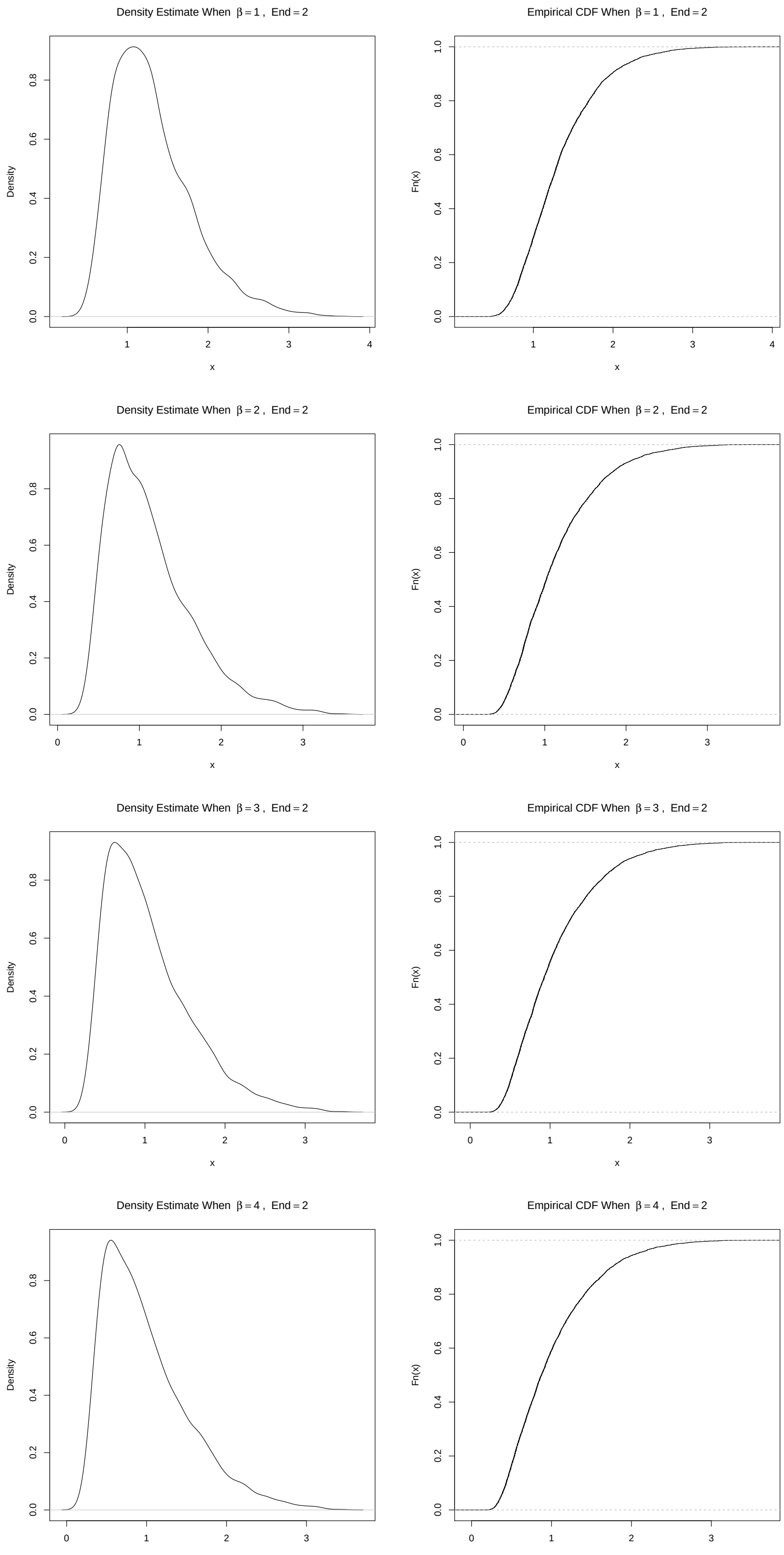


Figure 4.4. Simulated density and distribution function of $\Theta(3)$

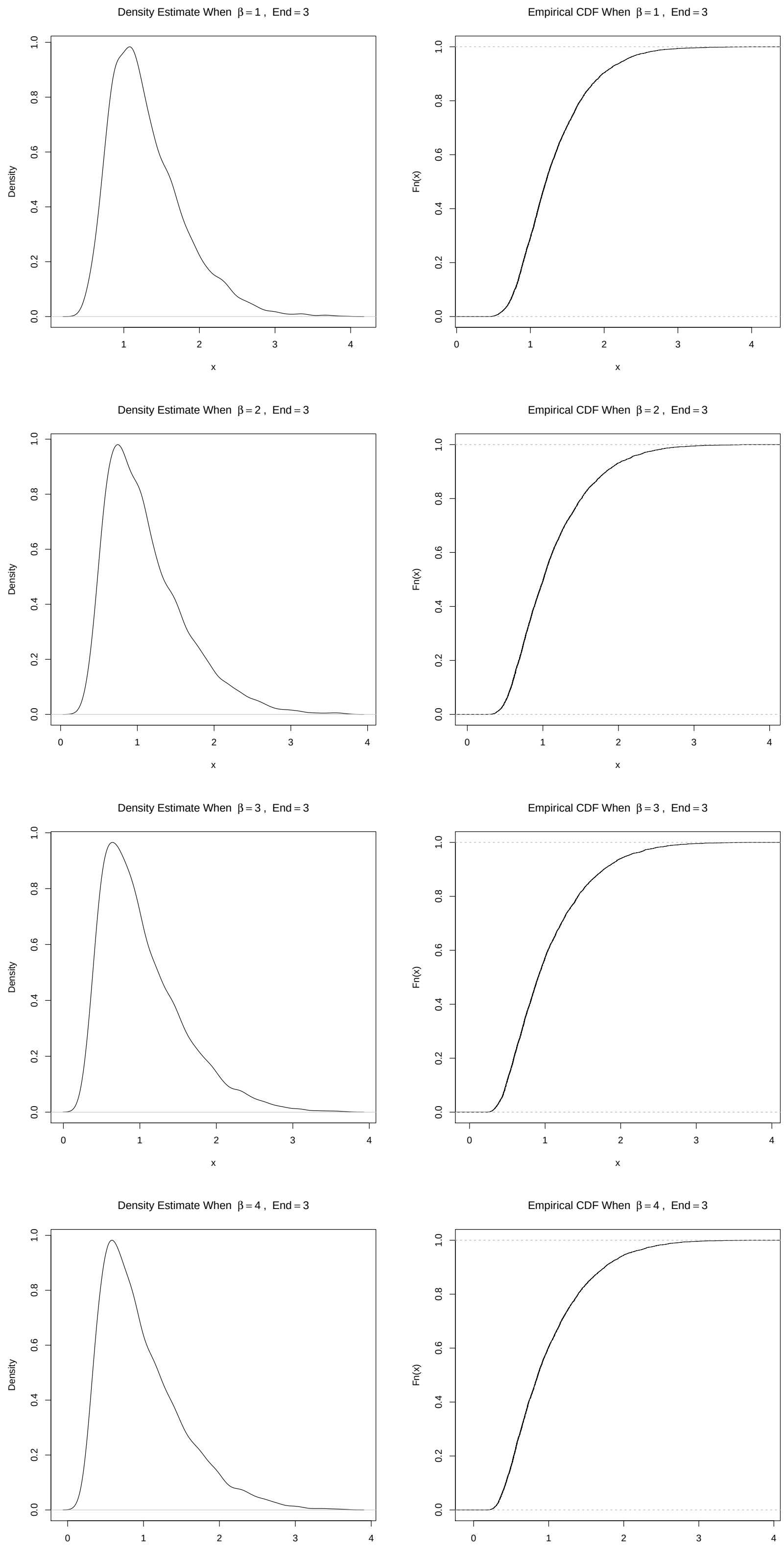


FIGURE 4.5. Simulated density and distribution function of $\Theta(4)$

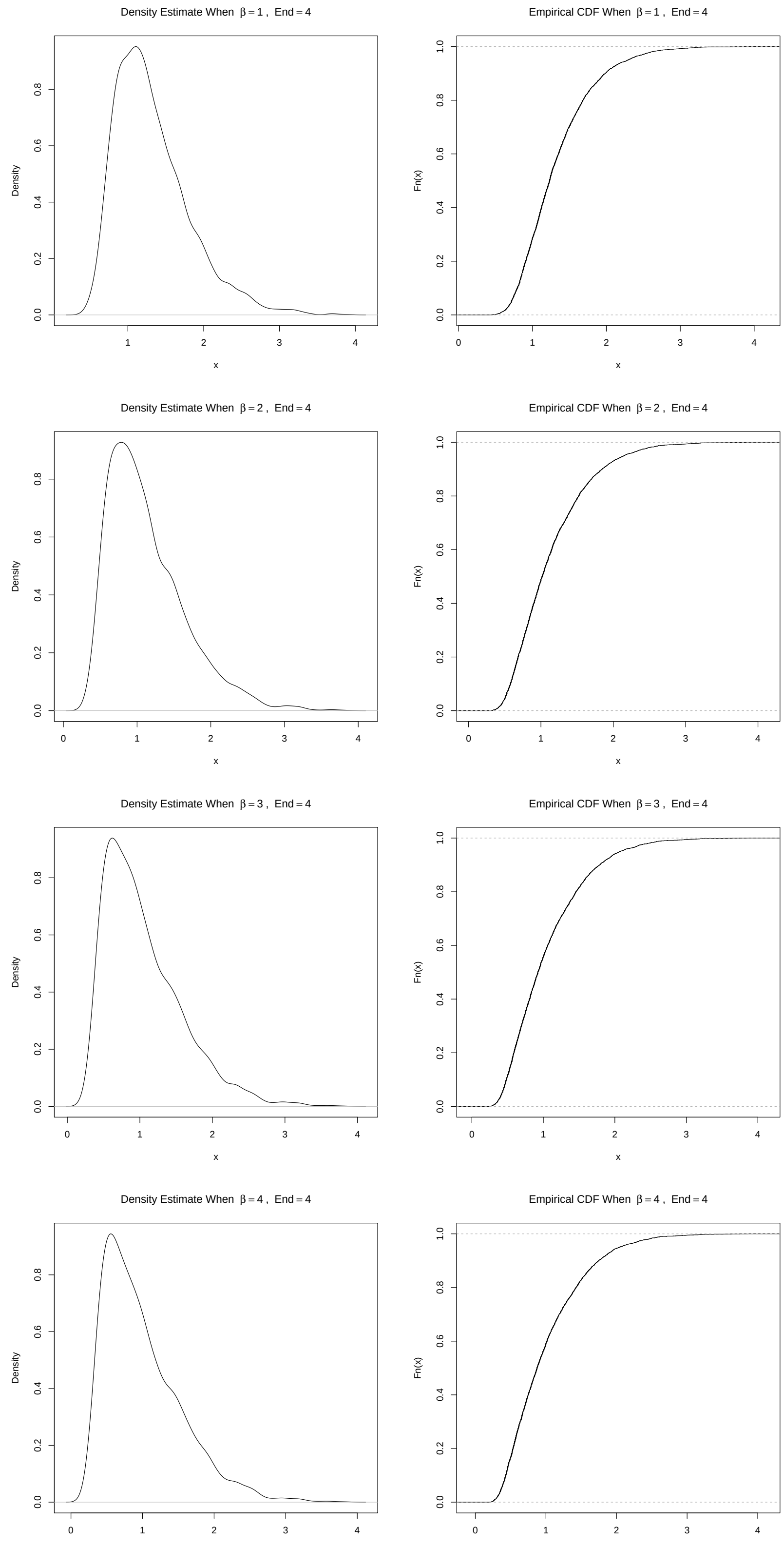


FIGURE 4.6. Simulated density and distribution function of $\Theta(5)$

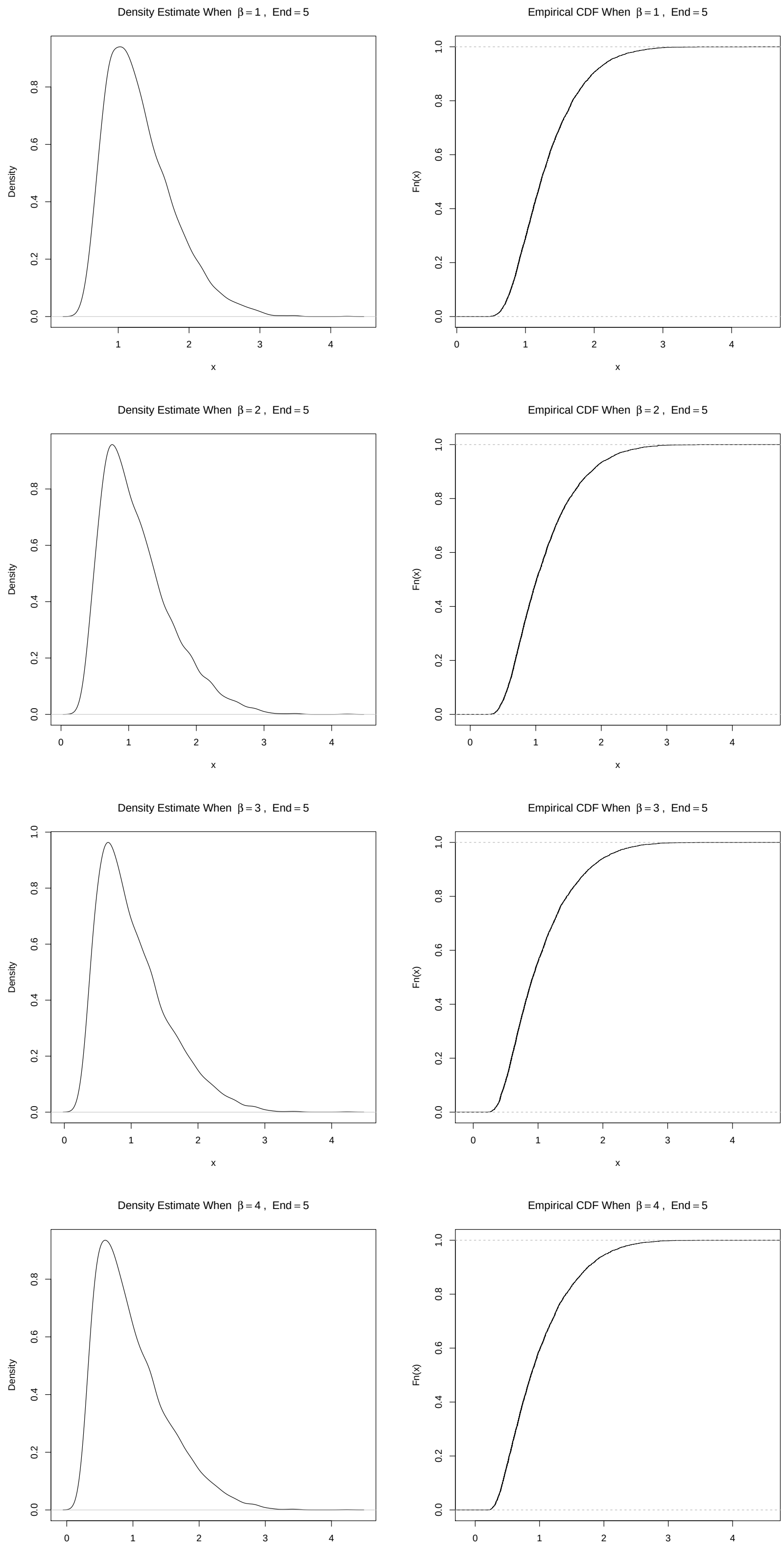


Figure 4.7. Simulated density and distribution function of $\Theta(6)$

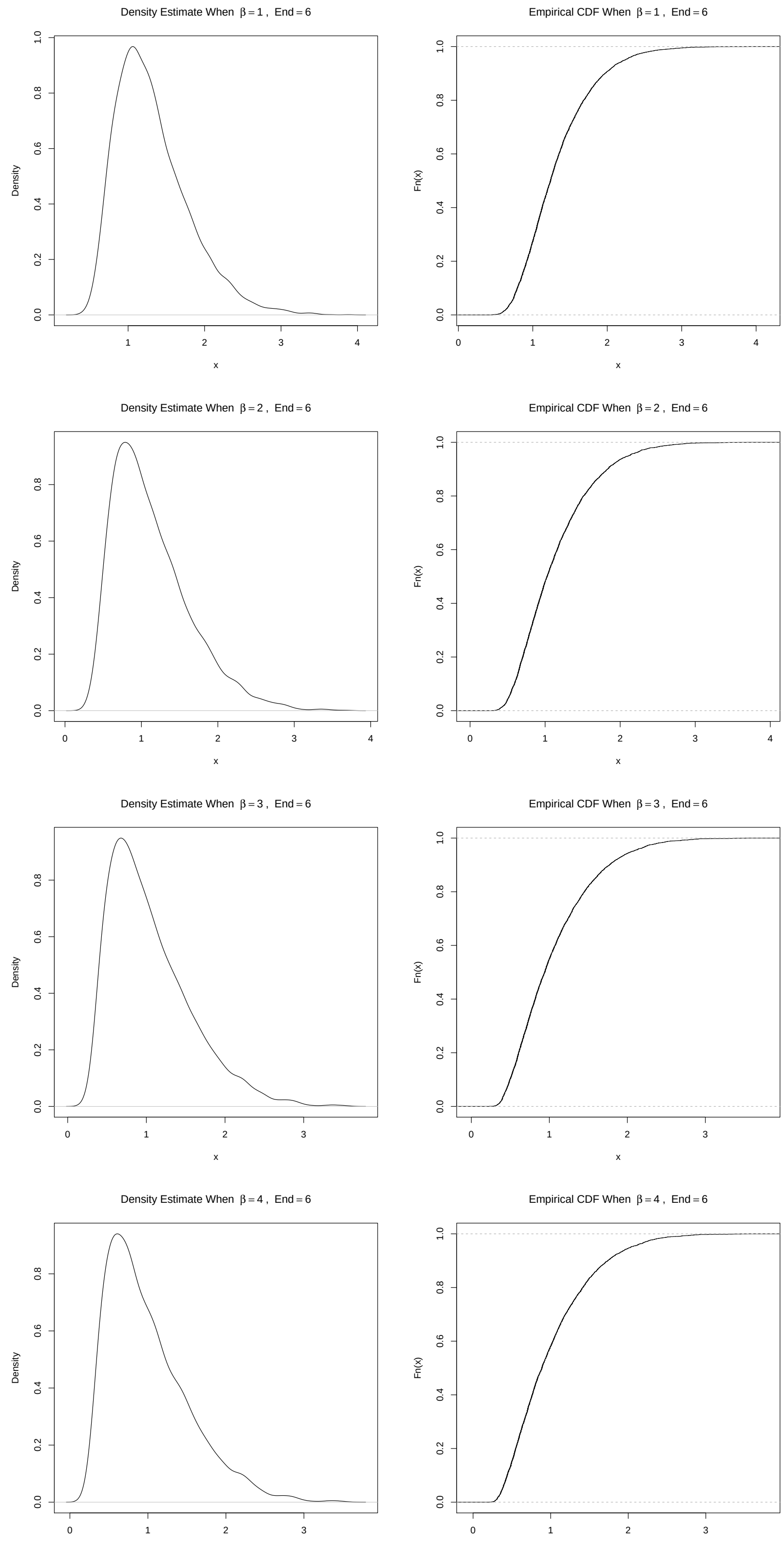


Table 4.1. Upper quantiles (critical values) for $\Theta(u)$

β	u	$c_{90\%}$	$c_{95\%}$	$c_{99\%}$
1	1	1.945051	2.199489	2.731077
	2	1.981252	2.266893	2.809544
	3	1.974247	2.275988	2.817490
	4	1.985353	2.291374	2.900883
	5	1.976551	2.231797	2.748361
	6	1.961971	2.236345	2.725577
	∞	1.961971	2.236345	2.725577
2	1	1.814392	2.095318	2.624788
	2	1.840018	2.145515	2.721245
	3	1.839820	2.150535	2.729268
	4	1.846423	2.148506	2.731067
	5	1.850786	2.128663	2.635953
	6	1.813968	2.105014	2.625778
	∞	1.813968	2.105014	2.625778
3	1	1.758939	2.043119	2.568253
	2	1.785773	2.083754	2.696876
	3	1.775451	2.080368	2.707124
	4	1.790178	2.070710	2.685414
	5	1.775937	2.073220	2.585892
	6	1.756301	2.046612	2.586148
	∞	1.756301	2.046612	2.586148
4	1	1.724598	2.022846	2.552377
	2	1.749547	2.061546	2.675776
	3	1.743404	2.045160	2.683522
	4	1.747410	2.041674	2.632820
	5	1.748103	2.046688	2.571845
	6	1.735013	2.010628	2.568760
	∞	1.735013	2.010628	2.568760

Table 4.2. Asymptotic critical values chosen for Θ with different values of β

Critical Values			
β	$c_{90\%}$	$c_{95\%}$	$c_{99\%}$
1	1.961971	2.236345	2.725577
2	1.813968	2.105014	2.625778
3	1.756301	2.046612	2.586148
4	1.735013	2.010628	2.568760

5. Test Size for Independent Identically Distributed Random Varianbles using $\Theta_{M,T}$

In the following simulations, σ is estimated to be the square root of the sample variance. The variables are generated independently and $h = \lfloor T^{1/2} \rfloor$ which satisfies the assumptions needed for definitions (3.1) and (3.2). In all these simulations we are seeking the empirical size of the test to be close to the selected value of α. We define α as

$$\alpha = \lim_{M\to\infty} \mathrm{P}_{H_0}\left(\Theta_{M,T} \geq c_{1-\alpha}\right).$$

The random variables $\{X_i, \quad 0 \leq i \leq M+T\}$ are generated from the distribution listed in the table title.

TABLE 5.1. Empirical performance of test size, where $X_i \sim N(0,1)$

			$c_{.90}$		$c_{.95}$		$c_{.99}$	
M	T	β	$\mathrm{P}(\tau_M < T-h)$	$\mathrm{P}(\Theta_{M,T} > c)$	$\mathrm{P}(\tau_M < T-h)$	$\mathrm{P}(\Theta_{M,T} > c)$	$\mathrm{P}(\tau_M < T-h)$	$\mathrm{P}(\Theta_{M,T} > c)$
100	100	1	0.0718	0.0718	0.0344	0.0344	0.0066	0.0066
		2	0.0514	0.514	0.023	0.023	0.0032	0.0032
		3	0.0352	0.0352	0.0124	0.0124	0.0012	0.0012
		4	0.0226	0.0226	0.0066	0.0066	0.0	0.0
1000	100	1	0.0536	0.0536	0.0216	0.0216	0.0042	0.0042
		2	0.0358	0.0358	0.0134	0.0134	0.0016	0.0016
		3	0.0216	0.0216	0.007	0.007	0.0	0.0
		4	0.0126	0.0126	0.0036	0.0036	0.0	0.0
100	1000	1	0.1312	0.1312	0.0698	0.0698	0.0244	0.0244
		2	0.1146	0.1146	0.0636	0.0636	0.019	0.019
		3	0.1038	0.1038	0.0564	0.0564	0.0148	0.0148
		4	0.092	0.092	0.0508	0.0508	0.0128	0.0128
1000	1000	1	0.0812	0.0812	0.0394	0.0394	0.0086	0.0086
		2	0.0758	0.0758	0.0356	0.0356	0.0066	0.0066
		3	0.0686	0.0686	0.0314	0.0314	0.0056	0.0056
		4	0.0612	0.0612	0.0272	0.0272	0.0042	0.0042

As we can see from the table above, the size of the test for the sequential procedure and the maximum is equivalent in the empirical results. The tables that follow only show the size of the test for $\Theta_{M,T}$ assuming that $\mathrm{P}_{H_0}(\Theta_{M,T} > c) = \mathrm{P}_{H_0}(\tau_M < T-h) = \alpha$.

TABLE 5.2. Empirical performance of test size, where $X_i \sim \text{Exp}(\lambda = 2)$

M	T	β	$\text{P}(\Theta_{M,T} > c_{.90})$	$\text{P}(\Theta_{M,T} > c_{.95})$	$\text{P}(\Theta_{M,T} > c_{.99})$
100	100	1	0.1	0.0688	0.033
		2	0.0774	0.0474	0.0192
		3	0.0612	0.0342	0.0122
		4	0.043	0.0258	0.0088
1000	100	1	0.0854	0.053	0.022
		2	0.0648	0.037	0.013
		3	0.0494	0.026	0.007
		4	0.0346	0.018	0.0046
100	1000	1	0.1476	0.091	0.0422
		2	0.1276	0.0802	0.0314
		3	0.118	0.0734	0.0274
		4	0.107	0.0646	0.0236
1000	1000	1	0.0896	0.0512	0.0172
		2	0.0812	0.0438	0.014
		3	0.074	0.0388	0.0116
		4	0.0676	0.0358	0.0094

TABLE 5.3. Empirical performance of test size, when $X_i \sim T(\nu = 5)$

M	T	β	$\text{P}(\Theta_{M,T} > c_{.90})$	$\text{P}(\Theta_{M,T} > c_{.95})$	$\text{P}(\Theta_{M,T} > c_{.99})$
100	100	1	0.0822	0.044	0.0158
		2	0.0292	0156	0.0048
		3	0.004	0.0024	0.0
		4	0	0	0
1000	100	1	0.0548	0.0292	0.0082
		2	0.0234	0.011	0.0026
		3	0.0016	0.0	0.0
		4	0	0	0
100	1000	1	0.129	0.077	0.0126
		2	0.0754	0.04	0.0132
		3	0.0412	0.0204	0.0052
		4	0.0144	0.0066	0
1000	1000	1	0.0854	0.0436	0.0126
		2	0.0514	0.0242	0.0068
		3	0.0274	0.0116	0.0026
		4	0.0104	0.0038	0.0012

Table 5.4. Empirical performance of test size, when $X_i \sim \chi^2_{(\nu=5)}$

M	T	β	$P(\Theta_{M,T} > c_{.90})$	$P(\Theta_{M,T} > c_{.95})$	$P(\Theta_{M,T} > c_{.99})$
100	100	1	0.1158	0.0796	0.0322
		2	0.0902	0.0544	0.0196
		3	0.0696	0.039	0.0132
		4	0.0502	0.0254	0.0076
1000	100	1	0.1034	0.0636	0.0214
		2	0.0798	0.0412	0.0136
		3	0.0576	0.0284	0.0076
		4	0.0406	0.019	0.0034
100	1000	1	0.1546	0.0956	0.0156
		2	0.1396	0.0818	0.0328
		3	0.1278	0.0756	0.0304
		4	0.113	0.0688	0.0262
1000	1000	1	0.095	0.048	0.0156
		2	0.0826	0.0428	0.0122
		3	0.074	0.0398	0.001
		4	0.0656	0.0366	0.0082

Table 5.5. Empirical performance of test size, when $X_i \sim F_{(\nu_1=6,\nu_2=7)}$

M	T	β	$P(\Theta_{M,T} > c_{.90})$	$P(\Theta_{M,T} > c_{.95})$	$P(\Theta_{M,T} > c_{.99})$
100	100	1	0.1382	0.104	0.0622
		2	0.0712	0.047	0.0242
		3	0.0228	0.0136	0.007
		4	0.0036	0.0024	0.001
1000	100	1	0.0914	0.0638	0.036
		2	0.0454	0.0318	0.0168
		3	0.0164	0.0102	0.0048
		4	0.003	0.0016	0.0
100	1000	1	0.1878	0.1408	0.0372
		2	0.1146	0.086	0.0496
		3	0.0784	0.0.0552	0.031
		4	0.0444	0.0282	0.0136
1000	1000	1	0.0986	0.0672	0.0372
		2	0.0726	0.0514	0.0254
		3	0.0506	0.0342	0.0176
		4	0.0286	0.0182	0.009

TABLE 5.6. Empirical performance of test size, when $X_i \sim \text{Unif}(0,5)$

M	T	β	$\text{P}(\Theta_{M,T} > c_{.90})$	$\text{P}(\Theta_{M,T} > c_{.95})$	$\text{P}(\Theta_{M,T} > c_{.90})$
100	100	1	0.1014	0.0572	0.0166
		2	0.077	0.04	0.0082
		3	0.0574	0.0262	0.0044
		4	0.0388	0.0158	0.002
1000	100	1	0.095	0.0498	0.0122
		2	0.0716	0.00338	0.0054
		3	0.0508	0.0222	0.003
		4	0.0332	0.012	0.001
100	1000	1	0.1482	0.0854	0.029
		2	0.134	0.075	0.0234
		3	0.1232	0.068	0.0186
		4	0.1126	0.0608	0.0144
1000	1000	1	0.1022	0.047	0.0136
		2	0.093	0.0448	0.0116
		3	0.0852	0.04	0.0094
		4	0.0752	0.0356	0.0078

TABLE 5.7. Empirical performance of test size, when $X_i \sim \text{Gamma}(k=1, \theta=1)$

M	T	β	$\text{P}(\Theta_{M,T} > c_{.90})$	$\text{P}(\Theta_{M,T} > c_{.95})$	$\text{P}(\Theta_{M,T} > c_{.99})$
100	100	1	0.1114	0.074	0.0376
		2	0.081	0.055	0.0248
		3	0.0652	0.0406	0.0142
		4	0.0508	0.0296	0.0082
1000	100	1	0.0758	0.045	0.0176
		2	0.0798	0.0412	0.0136
		3	0.0402	0.0198	0.0046
		4	0.0282	0.0142	0.0022
100	1000	1	0.1466	0.1002	0.016
		2	0.1258	0.0842	0.0378
		3	0.1168	0.0764	0.0318
		4	0.107	0.0718	0.0266
1000	1000	1	0.0908	0.05	0.016
		2	0.0796	0.0442	0.0116
		3	0.073	0.0384	0.0084
		4	0.0658	0.034	0.007

Based on the results shown for i.i.d. random variables, it is clear to see that when $\beta = 1$ we get empirical results that are the closest to the chosen test size α. The above examples also demonstrate the consistency of the test regardless of what type of distribution is being simulated from. We verified that for independent random variables this test does not need assumptions on the distribution of the random variable. We merely need stationarity and conditions on the second moment to ensure a finite variance which is consistent with our assumptions thus far. To further demonstrate the capabilities of the test in the i.i.d. case, we simulated τ_M under H_A. According to (2.20) we should get that the sequential procedure stops at a change point with approximately probability converging to one. Table 5.8 shows that this is indeed the case. The first examples shown in Table 5.8 increment the mean away from the training sample mean in order to get a sense for how the numerical difference of the mean influences the change detection. We also used other types of distributions with means away from zero to verify this consistency. The first example is a good indicator of how the values of β influence how conservative the test is when the difference between the hypothesized mean and the true mean is small.

TABLE 5.8. Sample stopping times under H_A, where training sample is N(0,1), $\alpha = 0.1, M = T = 100$

Test Sample	β	P($\tau_M < T - h$)	Test Sample	P($\tau_M < T - h$)	Test Sample	P($\tau_M < T - h$)
N(1,1)	1	0.9488	N(2,1)	1	N(3,1)	1
	2	0.8942		1		1
	3	0.855		1		1
	4	0.8072		1		1
N(4,1)	1	1	$\chi^2_{(2)}$	1	Gamma($k = 5, \theta = 1/5$)	0.9932
	2	1		0.9994		0.9908
	3	1		0.9988		0.9802
	4	1		0.997		0.9510
Exp($\lambda = 1/5$)	1	1	Unif(0, 5)	1	Poisson($\mu = 1$)	0.9422
	2	1		1		0.8960
	3	1		1		0.8948
	4	1		1		0.8004

6. Dependent Sequences and Long-Run Variance

The previous cases have used i.i.d. random variables to obtain the desired results of the testing procedure presented in Section 2. The following material presents the necessary conditions to apply this procedure when there is a dependence among the sequence of random variables. There are various dependent process models that we would like to evaluate using $\Theta_{M,T}$ with the correct critical values and the appropriate value of σ for the boundary function. We show that the critical values found in Table 4.2 are the same for these settings and that the estimator for σ has a simple closed form. Before theorems are presented and the models are analyzed there is a need for a few definitions and some clarification of notation. This notation is used throughout the derivation of the long-run cumulative impulse response or long-run variance as we refer to it here, denoted by $\sigma^2_{x_t}$.

The goal is to estimate the long-run variance for a particular process in a non-parametric way. The the parametric form of the long-run variance is derived using a few well-known theorems and which then leads to the non-parametric estimator. We use Wold's Decomposition and the infinite order auto regressive moving average representation we showed in Section 2 to arrive at the parametric form when the model parameters are known.

The main component of the long-run variance is the autocovariance function. The autocovariance function is denoted by

$$\gamma_j = \mathrm{Cov}(X_t, X_{t+j}). \tag{6.1}$$

White noise is assumed to be normally distributed with mean zero and variance σ_w^2, i.e.

$$w_j \sim N(0, \sigma_w^2). \tag{6.2}$$

Unless otherwise specified in simulations, we assume that $\sigma_w^2 = 1$. The sample autocovariance function is denoted by

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} \left(x_{t+|h|} - \bar{x}\right)\left(x_t - \bar{x}\right) \text{ for } -n < h < n. \tag{6.3}$$

Wold's decomposition theorem (Fuller (1996)) states that any covariance stationary time series X_t has a linear process or infinite order moving average representation of the form

$$X_t = \mu + \sum_{k=0}^{\infty} \psi_k w_{t-j},$$

where the weights ψ_j are assumed to have the following properties

$$\psi_0 = 1, \sum_{k=0}^{\infty} \psi_k < \infty.$$

In Wold form it can be shown that

$$\begin{aligned}
\mathrm{E}[X_t] &= \mu, \\
\gamma_0 = \mathrm{Var}(X_t) &= \sigma_w^2 \sum_{k=0}^{\infty} \psi_k^2, \\
\gamma_j = \mathrm{Cov}(X_t, X_{t-j}) &= \sigma_w^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+j}.
\end{aligned}$$

For a stationary and ergodic time series $\lim_{s \to \infty} \psi_s = 0$ and the long-run cumulative impulse response $\sum_{s=0}^{\infty} \psi_s < \infty$. The Wold representation in lag operator notation is

$$\psi(L) = \sum_{k=0}^{\infty} \psi_k L^k, \quad \psi_0 = 1.$$

With ARMA(p, q) models the Wold polynomial is approximated by the ratio of the AR and MA polynomials when the assumptions of causality are met or when the roots of $\delta(z)$ and $\theta(z)$ lie outside the unit circle

$$\psi(1) = \frac{\theta(1)}{\delta(1)}.$$

For GARCH models, instead of the Wold representation, the unconditional variance of the series is used. It is calculated by the ratio of the mean to the ARCH and GARCH component polynomials $\alpha(L)$ and $\beta(L)$ respectively

$$\psi(1)^2 = \begin{cases} \dfrac{\omega}{\alpha(1)} & \text{ARCH}(p), \\ \dfrac{\omega}{\alpha(1) + \beta(1) - 1} & \text{GARCH}(p, q). \end{cases}$$

We rely on Theorem 6.1 given below to keep our analysis consistent with the convergence we established for $\Theta_{M,T}$ with independent random variables. The form of the long-run variance is due to Newman and Wright, (1981). The following definition is an important assumption in the following theorems.

Definition 6.1. *Two random variables X, Y are associated if* $\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$ *is non-negative.*

Theorem 6.1. *(Newman and Wright, 1981) Suppose $X_1, X_2, ...$ is a non-degenerate, strictly stationary, finite variance sequence which is associated and such that*

$$\sigma^2 = \mathrm{Cov}(X_1, X_1) + 2\sum_{j=2}^{\infty} \mathrm{Cov}(X_1, X_j) < \infty. \tag{6.4}$$

For each $n = 1, 2, ...,$ define the stochastic process

$$\begin{aligned} W_n(t) &= [X_1 + \cdots + X_m + (nt - m)X_{m+1} - ntE(X_1)]/(\sigma\sqrt{n}) \\ &\quad \textit{with } m/n \leq t < (m+1)/n \end{aligned} \tag{6.5}$$

for $0 \leq t \leq T - h$; then the sequence of processes W_n converges in distribution (in $C[0, T]$) to the standard Wiener process.

Theorem 6.2. *Let $\{X_j : j \in \mathbb{N}\}$ be a strictly stationary sequence of associated random variables with $\mathrm{E}[X_j] = 0, \mathrm{E}[X_j^2] < \infty$. Assume*

$$0 < \sigma^2 = \mathrm{Cov}(X_1, X_1) + 2\sum_{j=2}^{\infty} \mathrm{Cov}(X_1, X_j) < \infty \tag{6.6}$$

Then $\{X_j : j \in \mathbb{N}\}$ fulfulls the invariance principle.

Our forthcoming definition of long-run variance coincides with the theorems given above. Note that even if we don't have stationarity, we can still get the invariance principle with other conditions on the random variables. Consider the following Theorem,

Theorem 6.3. *Let $\{X_j : j \in \mathbb{N}\}$ be a sequence of associated random variables with $\mathrm{E}[X_j] = 0, \mathrm{E}[X_j^2] < \infty$. Assume*

$$\sigma_n^{-2}\mathrm{E}[S_{nk}S_{nl}] \xrightarrow{n} \min\{k, l\}, \tag{6.7}$$

$$\left\{\sigma_n^{-2}(S_{m+n} - S_m)^2 : m \in \mathbb{N} \cup \{0\}, n \in \mathbb{N}\right\} \textit{ is uniformly integrable.} \tag{6.8}$$

Then $\{X_j : j \in \mathbb{N}\}$ fulfills the invariance principle.

These are now conditions on the expectation of the partial sums which do not include stationarity. These assumptions along with the assumptions in Section 2 give an opportunity to generalize what has been found in our results. In general we are assuming that the limiting distribution of these random variables is not a constant. Thus we can reasonably assume that they are non-degenerate. We also not that we need a finite variance to work for the same theorem, however, this is already necessary because recall in the function $g(h,k)$ we have σ in the denominator.

In order for our test statistic, $\Theta_{M,T}$ to be valid according to the derivation, we need an estimator of the variance that is given by (6.6). We will use the methodology of the Long-Run Variance, Long-Run Cumulative Impulse Response, or Asymptotic Variance of the series X_t to obtain this result. Let X_t be a stationary and ergodic time series. Anderson's central limit theorem for stationary and ergodic processes (Hamilton (1994)) states

$$\sqrt{T}(\overline{X} - \mu) \xrightarrow{\mathcal{D}} N\left(0, \sum_{j=-\infty}^{\infty} \gamma_j\right),$$

or

$$\overline{X} \approx N\left(\mu, \frac{1}{T}\sum_{j=-\infty}^{\infty} \gamma_j\right).$$

The sample size, T, times the asymptotic variance of the sample mean is often called the long-run variance of x_t,

$$\sigma^2_{x_t} = \lim_{T\to\infty} T \operatorname{Var}(\overline{X}) = \sum_{j=-\infty}^{\infty} \gamma_j.$$

Since $\gamma_{-j} = \gamma_j$, $\sigma^2_{x_t}$ may be alternatively expressed as

$$\sigma^2_{x_t} = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j.$$

Using the Wold decomposition above, this can be rewritten further as

$$\sigma^2_{x_t} = \sigma^2_w \sum_{k=0}^{\infty} \psi_k^2 + 2\sum_{j=1}^{\infty} \gamma_j = \sigma^2_w \sum_{k=0}^{\infty} \psi_k^2 + 2\sigma^2_w \sum_{j=1}^{\infty}\left(\sum_{k=0}^{\infty} \psi_k\psi_{k+j}\right) = \sigma^2_w \left(\sum_{k=0}^{\infty} \psi_k\right)^2,$$

and again using the Wold representation, and what we derived for (2.14) we have

$$\sigma^2_{x_t} = \sigma^2_w \psi(1)^2. \tag{6.9}$$

We use (6.9) to derive the parametric long-run variance for the models used in simulation. A consistent estimate of $\sigma^2_{x_t}$ may be computed by first estimating the appropriate parameters of the chosen ARMA(p,q) model and then substituting that into (6.9). Note that a consistent estimate of $\sigma^2_{x_t}$ may also be computed using some non-parametric methods. An estimator made popular by Newey and West (1987), and the one we use for our examples is the weighted auto covariance estimator given by

$$\hat{\sigma}^2_{x_t} = \hat{\gamma}_0 + 2\sum_{j=1}^{R_M} w_{j,M}\hat{\gamma}_j, \tag{6.10}$$

where $w_{j,M}$ are weights which sum to unity and R_M is a truncation lag parameter that satisfies $R_M = O(M^{1/3})$. It was suggested by Newey and West that for general linear processes to use the Bartlett weights $w_{j,M} = 1 - j/(f(M)+1)$ where $f(M) = \lfloor 4(M/100)^{2/9} \rfloor$. In the following sections we use other well known kernels to evaluate if there are better choices of kernels which yield more consistent results for the monitoring procedure. We use $z = j/(f(M)+1)$ and M equal to the number of training examples in the kernel functions. These are all easily implemented within the simulations, and the definition of each kernel used is given in Table 6.1. The simulations performed on ARCH and GARCH models leave out the kernel weights due to the fact that these processes are uncorrelated as shown in Appendix C.

Table 6.1. Kernels used

Kernel Name	Abbreviation	Kernel, $K(z)$, where $0 \le z \le 1$
Bartlett	B.	$K(z) = 1 - z$
Parzen	P.	$K(z) = \begin{cases} 1 - 6z^2 + 6z^3, & 0 \le z \le 0.5, \\ 2(1-z)^3, & 0.5 < z \le 1, \end{cases}$
Flat-Top	F.T.	$K(z) = \begin{cases} 1, & 0 \le z \le 0.5, \\ 2(1-z), & 0.5 \le z \le 1, \end{cases}$
Quadratic Spectral	Q.S.	$K(z) = \frac{25}{12\pi^2 z^2}\left(\frac{\sin(6\pi z/5)}{6\pi z/5} - \cos(6\pi z)\right)$
Tukey-Hanning	T.H.	$K(z) = \frac{1+\cos(\pi z)}{2}$

7. Test Size for Dependent Random Variables using $\Theta_{M,T}$

For each model we evaluate, the parametric and non-parametric estimate are compared via simulation. The models evaluated are forms of ARMA(p,q) and GARCH(p,q) models that are the most widely used for dependent data. The first table for each model shows the empirical size of the test for the same sample size scenarios used in the previous simulations, using the parametric form of $\sigma^2_{x_t}$. The second table shows the empirical size of the test using the non-parametric long-run variance estimator with a fixed sample size in order to compare the kernels used. The goal with these simulations is to evaluate which values of β and which kernels give a more accurate size of the test. For the

AR(1) model, the parametric long-run variance we derived in (6.9) equates to

$$\sigma_{x_t}^2 = \sigma_w^2 \psi(1)^2 = \frac{\sigma_w^2}{\delta(1)^2} = \frac{\sigma_w^2}{(1-\delta_1)^2}.$$

In the following simulations, $\delta_1 = 0.2$ which satisfies the conditions for stationarity. Thus

$$\sigma_{x_t}^2 = \frac{1}{(1-0.2)^2} = 1.5625, \tag{7.1}$$

and $\sigma_{x_t} \approx 1.25$ in the boundary function for $\Theta_{M,T}$.

TABLE 7.1. Size of the test for AR(1) sequences with $\delta_1 = 0.2$ with given long-run variance

M	T	β	$P(\Theta_{M,T} > c_{.90})$	$P(\Theta_{M,T} > c_{.95})$	$P(\Theta_{M,T} > c_{.99})$
100	100	1	0.0588	0.029	0.008
		2	0.0496	0.202	0.0038
1000	100	1	0.0544	0.022	0.002
		2	0.0406	0.0158	0.002
100	1000	1	0.1176	0.057	0.02
		2	0.1062	0.0582	0 .0158
1000	1000	1	0.0734	0.039	0.0075
		2	0.0696	0.0328	0.0054

TABLE 7.2. Size of the test for AR(1) sequences with $\delta_1 = 0.2$ with estimated long-run variance

Kernel	β	$P(\Theta_{M,T} > c_{.90})$	$P(\Theta_{M,T} > c_{.95})$	$P(\Theta_{M,T} > c_{.99})$
B.	1	0.0908	0.0496	0.019
	2	0.0672	0.0348	0.0128
P.	1	0.1016	0.0568	0.022
	2	0.0738	0.0394	0.0138
F.T.	1	0.0776	0.041	0.0154
	2	0.0566	0.0304	0.0096
Q.S.	1	0.085	0.0442	0.0182
	2	0.0628	0.0326	0.0112
T.H.	1	0.0882	0.048	0.0186
	2	0.0646	0.034	0.0124

From these results we can see that the procedure is giving results that are consistent with the i.i.d. trials that were run. Another consistency we see is that when $\beta = 2$ we get empirical results that are closer to the true values of α. We did not include results when $\beta = 3$ or $\beta = 4$ because similar to the i.i.d. simulations, those values of β produce empirical Type 1 Error to be much lower than expected.

For the MA(1) model, the parametric long-run variance we derived in (6.9) equates to

$$\sigma_{x_t}^2 = \sigma_w^2 \psi(1)^2 = \sigma_w^2 \theta(1)^2 = (1+\theta_1)^2.$$

In the following simulations, $\theta_1 = 0.3$ which satisfies the conditions for stationarity. Thus $\sigma_{x_t} = \sqrt{(1+0.3)^2} = 1+0.3 = 1.7$ in the boundary function for $\Theta_{M,T}$.

TABLE 7.3. Size of the test for MA(1) sequences with $\theta_1 = 0.3$ with given long-run variance

M	T	β	$P(\Theta_{M,T} > c_{.90})$	$P(\Theta_{M,T} > c_{.95})$	$P(\Theta_{M,T} > c_{.99})$
100	100	1	0.0578	0.028	0.009
		2	0.039	0.0175	0.0025
1000	100	1	0.0446	0.002	0.06
		2	0.0325	0.0145	0.003
100	1000	1	0.116	0.05	0.007
		2	0.0925	0.049	0.0175
1000	1000	1	0.0718	0.034	0.0065
		2	0.064	0.032	0.005

TABLE 7.4. Size of the test for MA(1) sequences with $\theta_1 = 0.3$ with estimated long-run variance

Kernel	β	$P(\Theta_{M,T} > c_{.90})$	$P(\Theta_{M,T} > c_{.95})$	$P(\Theta_{M,T} > c_{.99})$
B.	1	0.1176	0.0286	0.0066
	2	0.0766	0.0396	0.0088
P.	1	0.1032	0.0566	0.0182
	2	0.0812	0.0446	0.0104
F.T.	1	0.1106	0.062	0.018
	2	0.0656	0.0312	0.008
Q.S.	1	0.0868	0.0462	0.0138
	2	0.0714	0.035	0.008
T.H.	1	0.0952	0.0522	0.0156
	2	0.0876	0.0502	0.0188

For the ARMA(1,1) model, the parametric long-run variance we derived in (6.9) equates to

$$\sigma_{x_t}^2 = \sigma_w^2 \psi(1)^2 = \sigma_w^2 \left(\frac{\theta(1)}{\delta(1)}\right)^2.$$

In the following simulations, $\delta_1 = 0.2, \theta_1 = 0.3$ which satisfies the conditions for stationarity. Thus

$$\sigma_{x_t}^2 = \left(\frac{1+0.3}{1-0.2}\right)^2,$$

and $\sigma_{x_t} = 1.625$ in the boundary function for $\Theta_{M,T}$.

TABLE 7.5. Size of the test for ARMA(1,1) sequences with $\delta_1 = 0.2, \theta_1 = 0.3$ with given long-run variance

M	T	β	$\mathrm{P}(\Theta_{M,T} > c_{.90})$	$\mathrm{P}(\Theta_{M,T} > c_{.95})$	$\mathrm{P}(\Theta_{M,T} > c_{.99})$
100	100	1	0.067	0.0328	0.008
		2	0.0514	0.0238	0.0052
1000	100	1	0.0684	0.0324	0.0074
		2	0.0522	0.0222	0.0038
100	1000	1	0.188	0.108	0.039
		2	0.1674	0.0986	0.031
1000	1000	1	0.1868	0.1104	0.039
		2	0.1694	0.1036	0.0332

TABLE 7.6. Size of the test for ARMA(1,1) sequences with $\delta_1 = 0.2, \theta_1 = 0.3$ with estimated long-run variance

Kernel	β	$\mathrm{P}(\Theta_{M,T} > c_{.90})$	$\mathrm{P}(\Theta_{M,T} > c_{.95})$	$\mathrm{P}(\Theta_{M,T} > c_{.99})$
B.	1	0.1096	0.0602	0.0226
	2	0.08	0.0446	0.014
P.	1	0.1362	0.0794	0.0294
	2	0.097	0.0552	0.0192
F.T.	1	0.0736	0.0384	0.0132
	2	0.0542	0.0288	0.0094
Q.S.	1	0.0912	0.0502	0.0182
	2	0.0668	0.0358	0.0116
T.H.	1	0.1036	0.0556	0.021
	2	0.0754	0.0418	0.0134

For the ARCH(1) model, the parametric long-run variance we obtain is taken from in (C.4). We first choose $\omega = 0.3$ and $\alpha_1 = 0.25$ to satisfy condition (C.6). This equates to

$$\sigma_{r_t}^2 = \frac{\omega}{1-\alpha_1} = \frac{0.3}{1-0.25} = 0.4.$$

Thus $\sigma_{r_t} \approx 0.632$ in the boundary function for $\Theta_{M,T}$. For the ARCH(p) process, we have by (C.2) that r_t is an uncorrelated process. As a result, the formulation we have for the long-run variance estimator becomes a simple calculation

$$\hat{\sigma}_{r_t}^2 = \hat{\gamma}_0.$$

Based on this estimator, we don't use kernels in the estimation of σ_{x_t} for the boundary function of $\Theta_{M,T}$.

TABLE 7.7. Size of the test for ARCH(1) sequences with $\omega = 0.3, \alpha_1 = 0.25$ with given long-run variance

M	T	β	$\text{P}(\Theta_{M,T} > c_{.90})$	$\text{P}(\Theta_{M,T} > c_{.95})$	$\text{P}(\Theta_{M,T} > c_{.99})$
100	100	1	0.0766	0.0342	0.0102
		2	0.0502	0.023	0.0064
1000	100	1	0.0558	0.0294	0.008
		2	0.0408	0.0176	0.0044
100	1000	1	0.119	0.068	0.021
		2	0.107	0.0582	0.0164
1000	1000	1	0.774	0.0386	0.0104
		2	0.0704	0.0342	0.0088

TABLE 7.8. Size of the test for ARCH(1) sequences with $\omega = 0.3, \alpha_1 = 0.25$ with estimated long-run variance

β	Kernel	$\text{P}(\Theta_{M,T} > c_{.90})$	$\text{P}(\Theta_{M,T} > c_{.95})$	$\text{P}(\Theta_{M,T} > c_{.99})$
1	None	0.0832	0.041	0.0134
2	None	0.0548	0.0258	0.007

For the GARCH(1,1) model, the parametric long-run variance we obtain is taken from in (C.4). We first choose $\omega = 0.3$ and $\alpha_1 = 0.25, \beta_1 = 0.1$ to satisfy condition (C.8). This equates to

$$\sigma_{r_t}^2 = \frac{\omega}{1 - \alpha_1 - \beta_1} = \frac{0.3}{1 - 0.25 - 0.1} \approx 0.462$$

Thus $\sigma_{r_t} \approx 0.679$ in the boundary function for $\Theta_{M,T}$.

TABLE 7.9. Size of the test for GARCH(1,1) sequences with $\omega = 0.3, \alpha_1 = 0.25, \beta_1 = 0.1$ with given long-run variance

M	T	β	$\text{P}(\Theta_{M,T} > c_{.90})$	$\text{P}(\Theta_{M,T} > c_{.95})$	$\text{P}(\Theta_{M,T} > c_{.99})$
100	100	1	0.0776	0.035	0.0108
		2	0.0506	0.0234	0.007
1000	100	1	0.0556	0.0292	0.0088
		2	0.0418	0.0196	0.0048
100	1000	1	0.1204	0.068	0.0206
		2	0.1086	0.058	0.0168
1000	1000	1	0.0786	0.0394	0.0106
		2	0.0708	0.035	0.0092

TABLE 7.10. Size of the test for GARCH(1,1) sequences with $\omega = 0.3, \alpha_1 = 0.25, \beta_1 = 0.1$ with estimated long-run variance

β	Kernel	$\mathrm{P}(\Theta_{M,T} > c_{.90})$	$\mathrm{P}(\Theta_{M,T} > c_{.95})$	$\mathrm{P}(\Theta_{M,T} > c_{.99})$
1	None	0.0842	0.0426	0.0138
2	None	0.0558	0.0272	0.0076

Based on the tables above we can see that when $\beta = 1$ we get closer to the specified size of the test. We conclude here that the best value of *beta* is 1 and the kernel that gives a test size closest to the chosen size is the Parzen kernel.

8. Visualization of Sequential Stopping Times

The sequential monitoring procedure was then applied to several processes where the training and test sets are simulated from models with different parameters to check the size of the test. Each training and test model is specified to satisfy the conditions of stationarity, but with different parameters to check how soon this procedure would detect the change. In each of these examples, $M = T = 100$, which gives the rolling window a value of $h = 10$, and we use the Parzen Kernel for the ARMA processes with the critical value chosen for $\alpha = 0.1$.

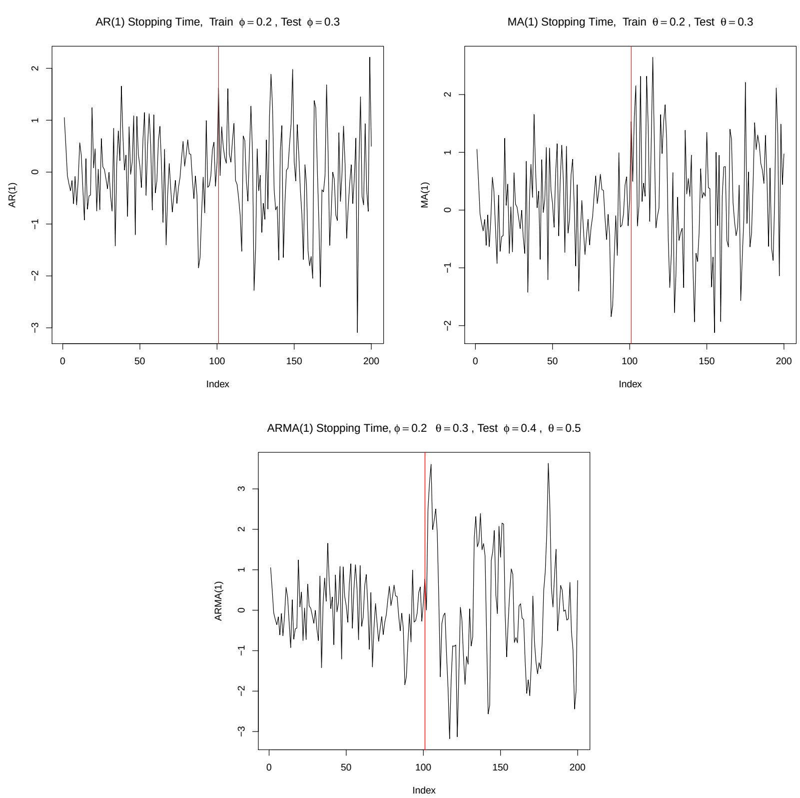


FIGURE 8.1. Simulated stopping time of ARMA processes

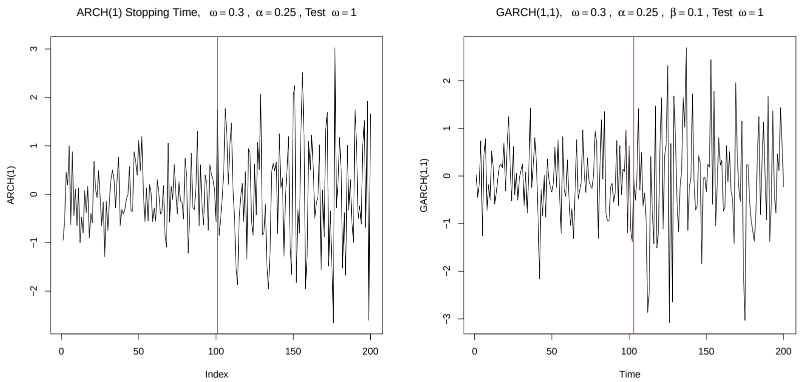


FIGURE 8.2. Simulated stopping time of ARCH(1), GARCH(1,1) processes

The plots above show, as expected, is that $\Theta_{M,T}$ is not just detecting the change, but quickly detecting the change in the mean or the parameter used to generate the test sample. In other examples that were simulated, some stopping times occurred later than others. However, in the parameterizations used above the sequential procedure always detected a change before the end of the test set. These tests would naturally lead into an analysis of the power function, however due to time constraints these are left as possibilities for future research. The main point here is that this test works when applied to simulated stationary dependent sequences with small changes in the mean or parameters. This synthetic example leads us to the main application of the testing procedure.

9. Applications

A natural application of the procedure presented is with financial and economic data. The application of these tests were performed on several stocks that historically have experienced extreme losses, bubbles, or have stayed neutral during an extreme economic event. More specifically we evaluate a handful of indexes from the great depression, Japan's bubble economy in the 1980s, the financial crisis of 2008, the "dot-com bubble" of 1999, and few that have recently experienced major shifts in market value. The groups are chosen from different scenarios to have applications in different time periods. Using the 'quantmod' package in R, we can look at the weekly, monthly, quarterly, and annual returns of these stocks and evaluate the procedure on detecting large deviations from normal return. As an introductory example, consider the Dow-Jones Industrial Average (DIJA) monthly share price in dollars (USD) between 1927 and 1934 obtained via FRED.

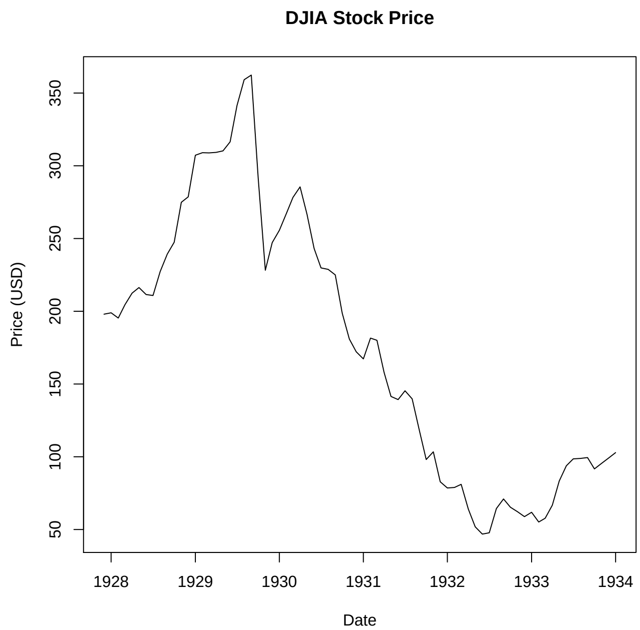


FIGURE 9.1. Dow-Jones Industrial Monthly Stock Price Index

In order to apply the test properly we need to convert the monthly price, S_t, into monthly log returns to have a better chance of meeting the stationarity assumption. We do so by taking the ratio of the monthly price at time t and the monthly return at time $t-1$ in order to calculate the ratio

$$\frac{S_t}{S_{t-1}},$$

and then we take the log of this ratio to get the return series

$$r_t = \log\left(\frac{S_t}{S_{t-1}}\right) = \log(S_t) - \log(S_{t-1}).$$

We then train the model on the first 12 months of the series.

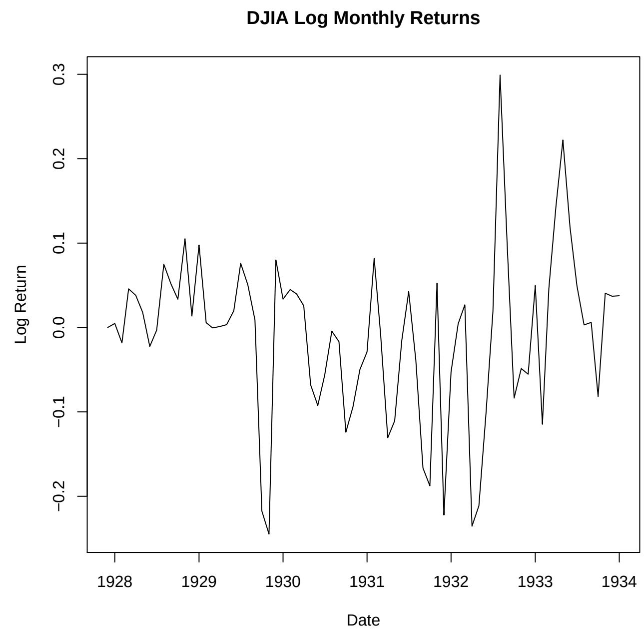
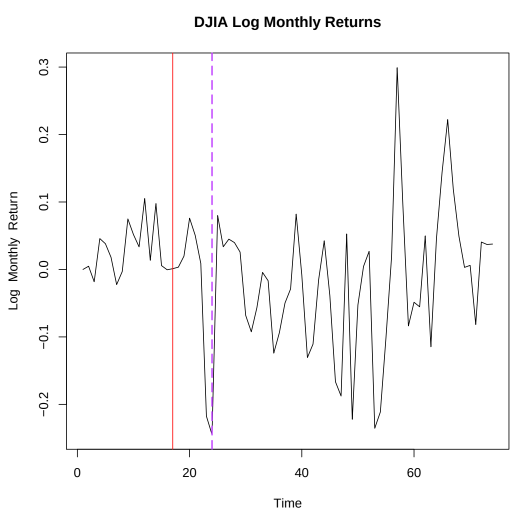


FIGURE 9.2. Dow-Jones Industrial log monthly stock returns

The Augmented Dickey-Fuller test is applied to the training sample and results in a p-value of 0.01 which is in favor of the series being stationary. We then run the sequential monitoring test on this series and obtain a stop time of April 1st, 1929. The window $h = 7$ which means that the moving average is comparing the returns between April and November of that year.



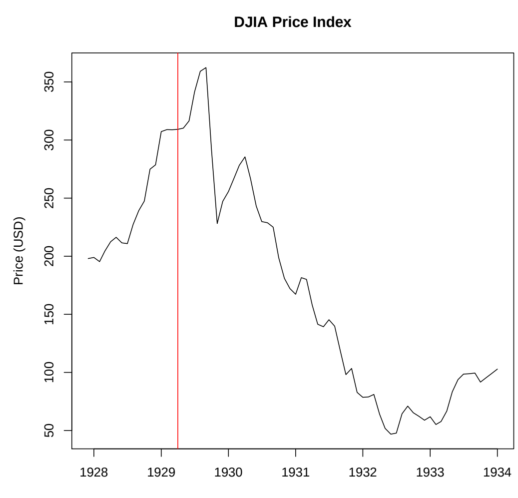


FIGURE 9.3. Dow-Jones Industrial log monthly stock returns and stop time

We can see from the test that the indicated stop time coincides with the period of time just before the sharp declines. This is a promising result that leads us to apply this to other stock prices and indices to gain a better understanding of how this test could be used in the context of stock returns.

To illustrate the capabilities of the test further we have three broader groups of securities with which to apply the test, namely 'loss', 'neutral', and 'gain'. These groups correspond to what type of change we are trying to detect. The 'loss' group will be stocks that lost significantly from the financial crisis of 2008 and a few more recent. These include American International Group (AIG), Bank of America (BAC), Citigroup (C), GoPro (GPRO), General Electric (GE), Goldman Sachs (GS), American Express (AXP), JP Morgan (JPM), and Zions Bank (Zion). These are among the hardest stocks hit during the 2008 financial crisis, AIG being the worst performing stock that year because of their insurance exposure to housing and housing derivative markets. The 'neutral' group is composed of stocks that weathered the declines of the 2008 financial crisis and tend to be regarded as 'recession proof' stocks. These include AutoZone (AZO), Clorox (CLX), Johnson and Johnson (JNJ), Dollar Tree (DLTR), and Wal-Mart (WMT). The 'gain' group is composed of stocks that experienced large bubbles before the dot-com crash of 1999 and Japan's bubble economy from the 1980's. This group includes Bitcoin (BTC), Cisco (CSCO), Microsoft (MSFT), Nikkei 225 Stock Index (N225) and Qualcomm (QCOM). These stock price histories are shown below as well as all test results are given in Appendix A. Each of these groups has various start dates which were chosen to be 1-3 years before the major change. This is to allow enough training data for the sample mean to be consistent and to meet the stationarity assumptions for the test. The abbreviations we use are summarized in Table 9.1 and the historical change dates are listed in Table A.2.

Table 9.1. Stock groups

Group Name	Abbreviations
Loss	AIG, AXP, BAC, C, GE, GS, JPM, ZION
Neutral	AZO, CLX, DLTR, JNJ, WMT
Gain (Bubble)	BTC, CSCO, GPRO, MSFT, N225, QCOM

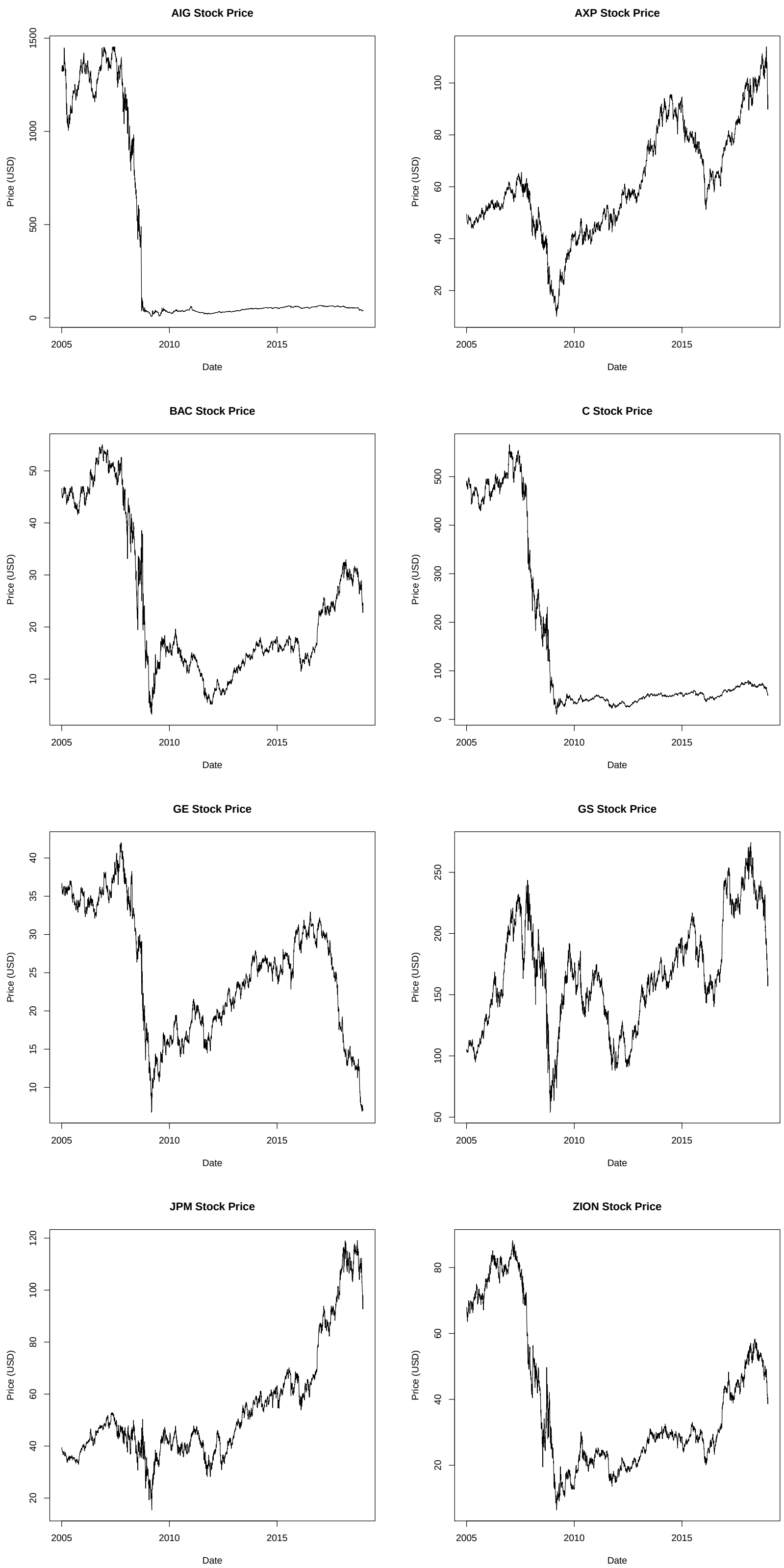


FIGURE 9.4. Historical stock prices of loss group

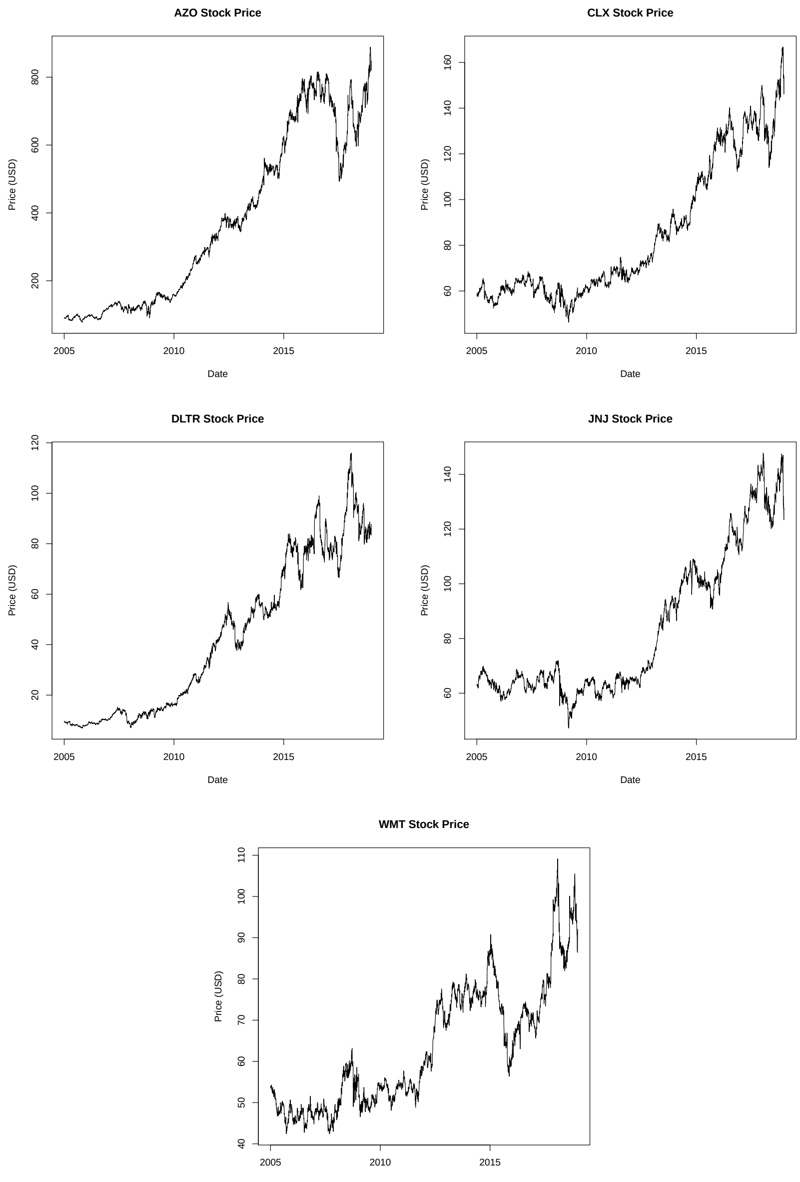


FIGURE 9.5. Historical stock prices of neutral group

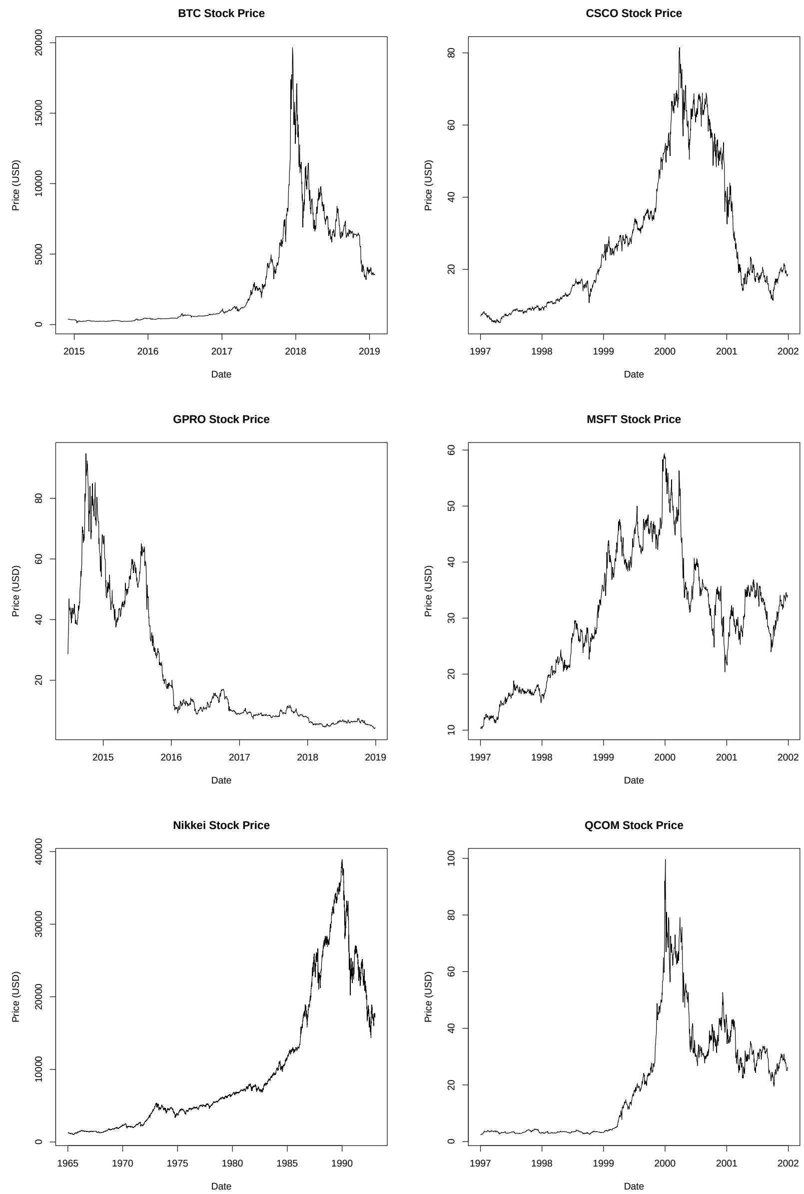


FIGURE 9.6. Historical stock prices of gain group

Before the test is applied, we need to evaluate which set among weekly, monthly, quarterly, and yearly returns would best meet the assumptions of the test and be the most practical. We start by visualizing the returns in each category for the stock and testing for stationarity. Weekly and monthly returns begin in the year 2005, and the other returns begin on the first trading day of the year 2000 to provide enough data in each category. Daily returns were ruled out due to the fact that they tend to have mean zero over the time periods examined even during more volatile periods in the actual closing price. The test would be possibly work better at smaller subsets of daily returns to apply the test and have significant change points but we leave this for future research. Sharp changes in the daily returns are difficult to detect with this procedure for the time period under consideration because they don't last long enough for a rolling moving average to pick up the change even with small h. For stationarity we use the Augmented Dickey-Fuller test on the return series for the training sample. The stocks we present as examples were checked to meet the stationarity assumption.

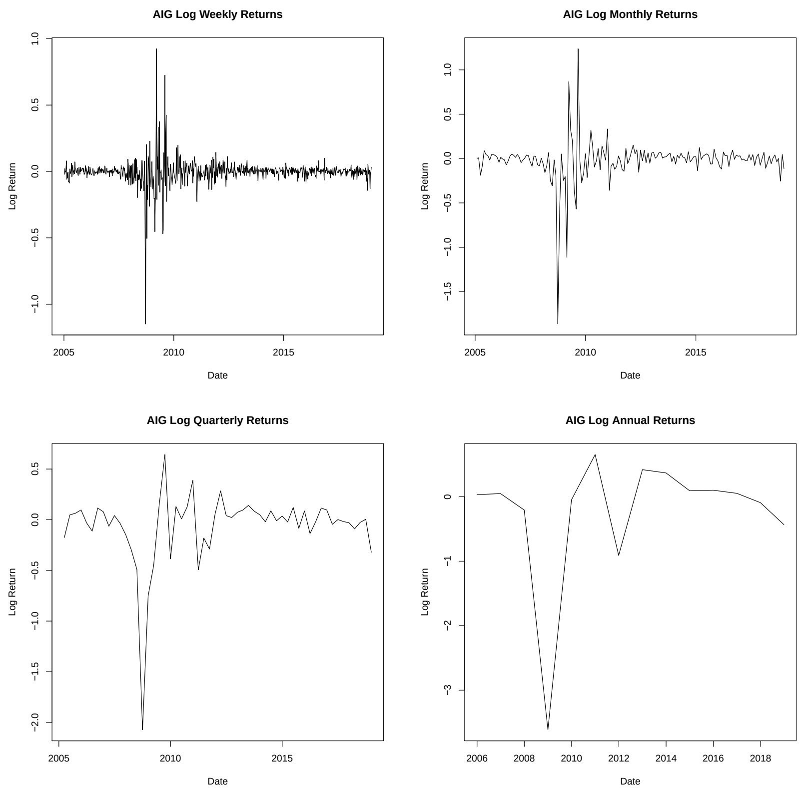


FIGURE 9.7. AIG stock returns

As we can see from these plots, the amount of data in each category as we move from weekly to annual returns become more limited. The choice was made to evaluate the test procedure on the weekly and monthly returns because of the amount of data available and the approximate stationarity of the first two years of data (2005 -2007 for AIG) in each series. In addition, the weekly and monthly returns capture more granular changes in the stock price that would be hidden when working with quarterly, and even annual returns. Another disadvantage of applying this test to quarterly returns is illustrated in the following scenario. The test was performed on the quarterly returns set of AIG and the resulting stop time in red is shown below.

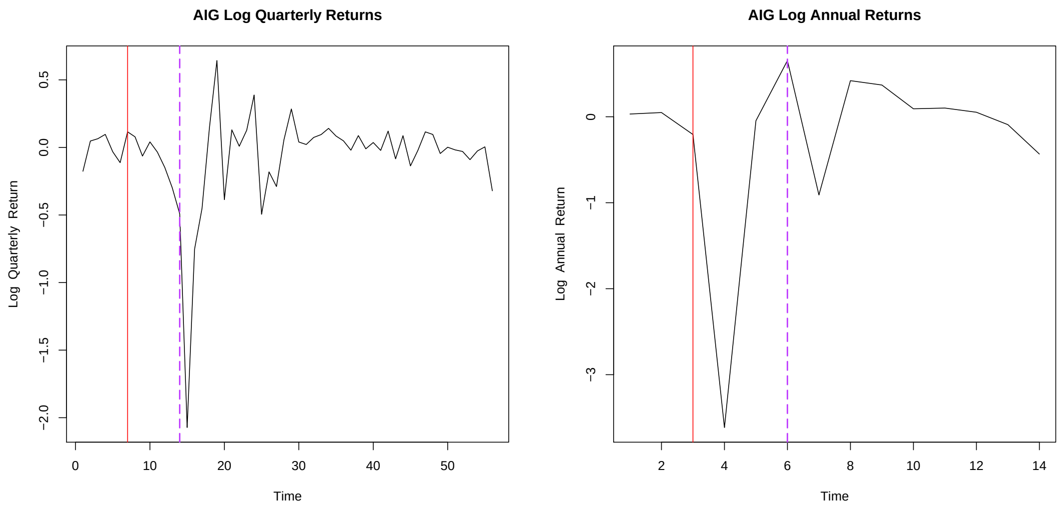


FIGURE 9.8. Quarterly stock return stop time

At first glance, the stop times seem to give a reasonable conclusion that if you were able to make a change or exit AIG on September 29th, 2006 or December 31st, 2007 of course you would have missed the large AIG price decline in 2008. However, in this case there are $T = 70$ points in the test set for quarterly returns and $T = 12$ for annual returns which imply $h = 8$ and $h = 3$ respectively. This means that the test has to look ahead three years for the annual returns and two years for quarterly returns. As the returns start to drift further apart, the test has to look further into future data to find a reasonable conclusion and becomes less practical for application. In the case of weekly returns, of the 730 weeks available between 2005 and 2018, we train the mean on the first 100 data points which leaves 630 data points for the test set. This gives us a training period from the beginning of 2005, through the middle of 2006. The test on weekly returns gives a rolling window of $h = 24$ which translates to 24 weeks, or about 6 months of future data. This seems more reasonable than the previous case. The results of the tests in weekly returns for a few of the loss stocks are shown below.

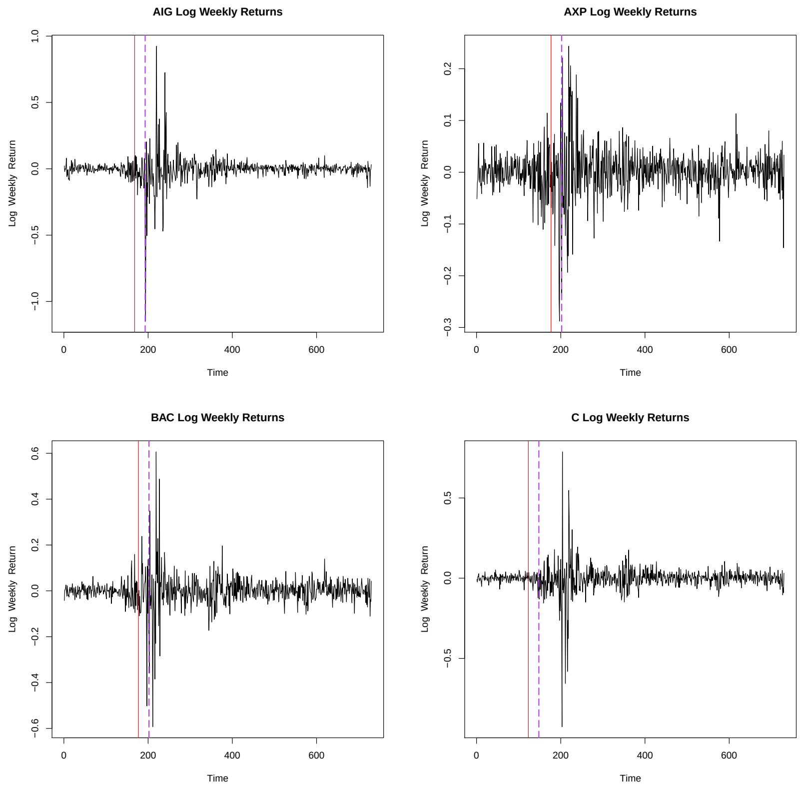


FIGURE 9.9. Weekly return stop times for loss group

From this sample of weekly stop dates, we can see that the method is picking up on significant changes to the stock returns. In the case of monthly returns, the rolling window stays around $h = 10$ which is 10 months into the future. This value of h is still a fairly reasonable window for the moving average. The stopping time of weekly and monthly returns gives a reasonable change point before the worst actually happens for these stock returns. Below are the closing prices of the change points discovered from both the monthly and weekly returns. Notice that the sequential monitoring procedure predicts the change before the bottom is actually reached indicating an acceptable point to exit the security.The stop times determined by the test on weekly and monthly returns are shown in each stock market history below. The weekly stop time is shown in red and the monthly stop time in blue.

TABLE 9.2. Legend

Week Stop Time
Month Stop Time
Price

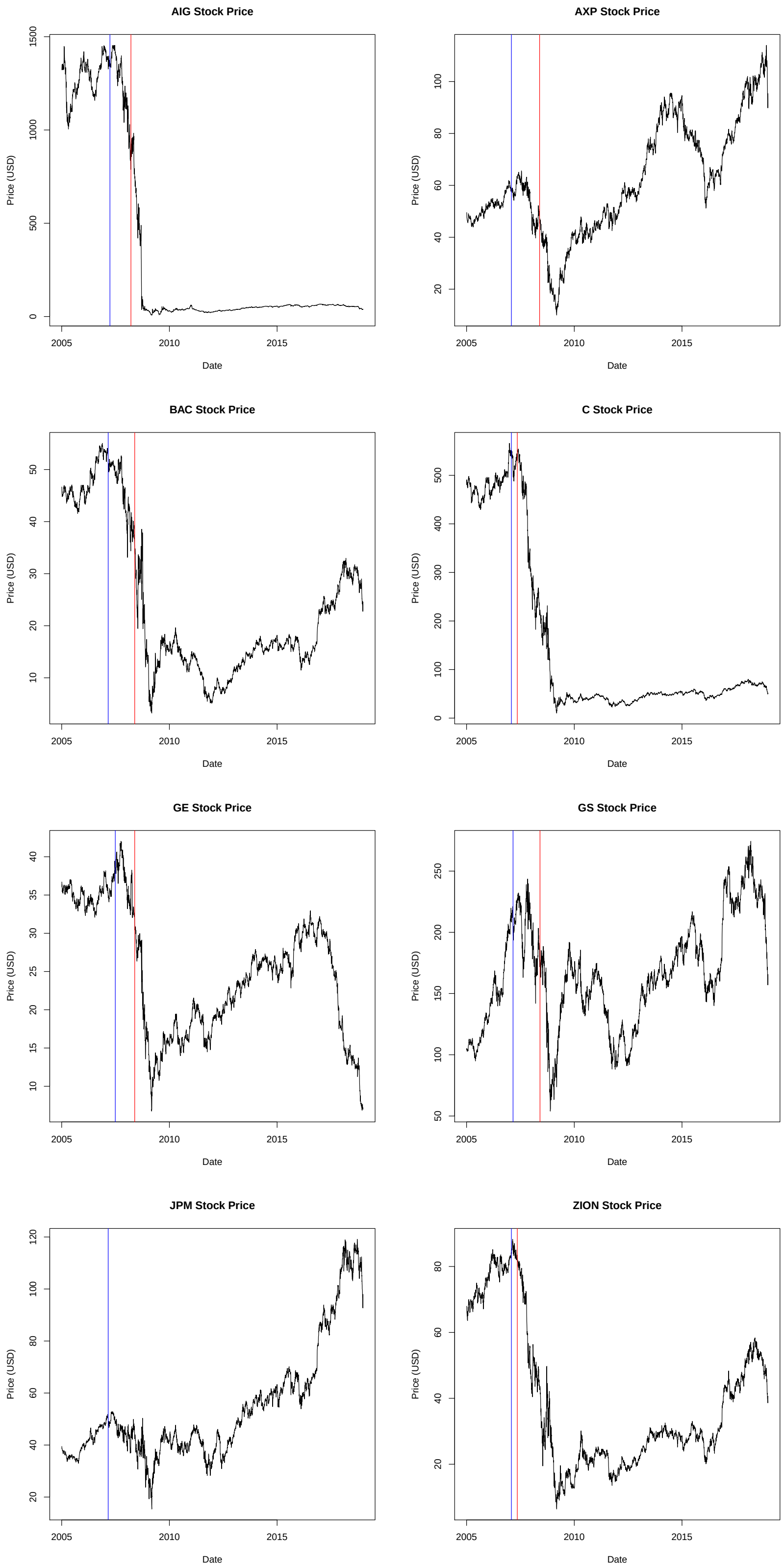


FIGURE 9.10. Daily closing price stop times for loss group

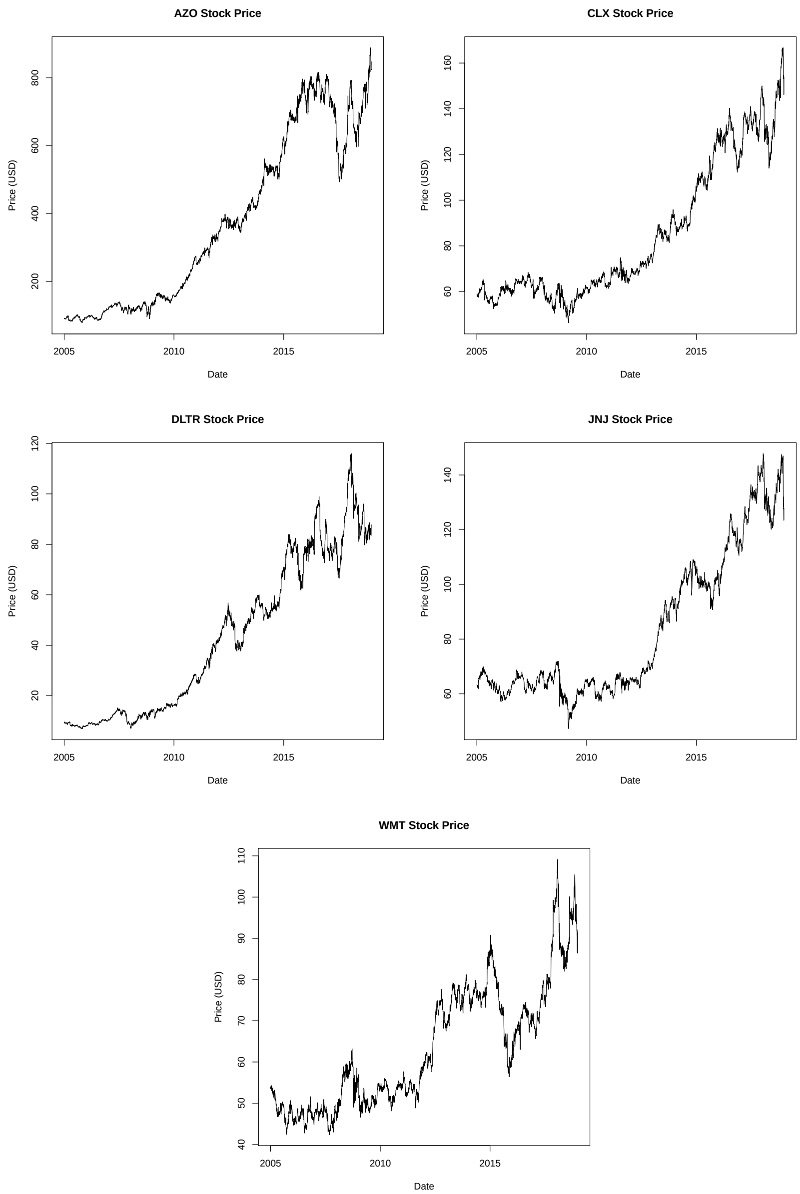


FIGURE 9.11. Daily closing price stop times for neutral group

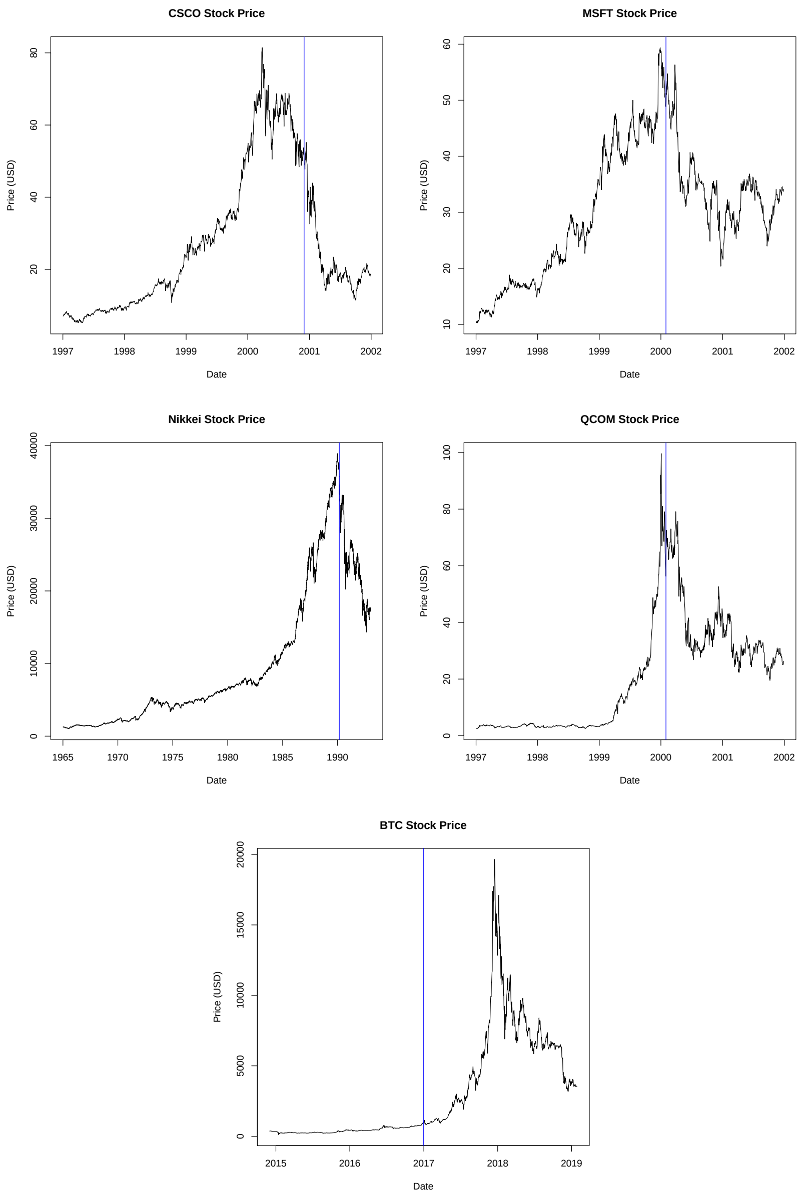


FIGURE 9.12. Daily closing price stop times for gain group

The weekly and monthly stop times found by the test consistently produce stopping times that come before a major decline. In the absence of a decline, the test appropriately does not signal a change or stopping point. For the gain or 'bubble' group that exhibits extreme appreciation in returns followed by steep declines, the test also appropriately signals a change before the decline reaches a bottom. There is a noticeable difference between the stopping times produced with weekly and monthly returns. Similar to daily returns, weekly returns seem to be too granular for some of these tests to pick up a major change point. This could be attributed to the amount of time it takes to reach a bottom for a decline, the steeper declines seem easier to detect with weekly returns. The longer declines seem to be signaled better by the monthly returns. These change points are also detected using the size of the test $\alpha = 0.1$, which means that if the size of the test is decreased it would be even harder for the weekly returns to detect a change.

10. Conclusion

From these simulations and applications we see that this sequential monitoring procedure can be used in practice to obtain change points that are significantly outside the typical distribution of examples observed until that point. This could be used for decision making processes with stocks, other time series data, and data that is assumed to be i.i.d.. The lack of computational complexity of the test allows it to be scalable to large data sets. The simulation run time was more heavily dependent on sampling from the particular model or distribution than the calculation of the test statistic. The simulations not only yielded a lot of theoretical results that can be used in practice, but also resulted in methods in R that make it simple to apply these procedures. We have a test that works well for i.i.d. observations, which could be used in place of other hypothesis testing procedures when others are not applicable because of assumption violations. The test works well for ARMA and GARCH processes, which could potentially extend to other types of time dependent models. The test has the possibility of extending more generally to processes that aren't stationary which would further the scope of applications.

The test we proposed is not without its limitations. When applied to real-time stock returns it is obviously not practical to look six to twelve months into the future. The rolling window can help or hinder the ability of the test to pick up less extreme changes in the distribution. For example, short-term decisions would be hindered if there is too much test data because the rolling window could be too large. The dependence on k in the boundary function which allows more emphasis on short term dynamics, hinders the test from picking up changes that are further away from the training sample. This could be remedied by updating the training sample if no significant change is found, checking for stationarity, and then continuing on with the test. However, these limitations only open opportunities for further research into other types of data that work well with this test. The simulations only gave a glimpse into the power of the test in each setting. Further research into the power function may also yield insight into what types of data and parametric assumptions are needed for the best results.

11. Acknowledgements

I thank my advisor, Lajos Horváth, for numerous draft revisions, direction, and guidance throughout this entire project. I thank all my former professors, especially Firas Rassoul-Agha, for helping me work through the proper simulation of Brownian motion and the random walk. I thank all my colleagues throughout my graduate career that have helped me work through many challenging problems. The problems we have worked through together prepared me to work through the large challenges associated with this project. I thank the University of Utah for giving me this incredible opportunity. I thank my family for being extremely supportive throughout my education and during this entire process. Finally I thank my wife, Amy, for always supporting me during the long hours spent putting this together.

Appendices

Appendix A. Stock Return Data

Table A.1. Stock stop time data when $\alpha = 0.1$

Stock	Weekly Returns						Monthly Returns					
	Train	M	T	h	$M+k^*$	Stop Date	Train	M	T	h	$M+k^*$	Stop Date
AIG	01/03/2005 - 12/01/2006	100	630	25	168	3/20/2008	01/31/2005 - 12/29/2006	24	144	12	27	03/30/2007
AXP	01/03/2005 - 12/01/2006	100	630	25	177	05/23/2008	01/31/2005 - 12/29/2006	24	144	12	25	01/31/2007
AZO	01/03/2005 - 12/01/2006	100	630	25	0	-	01/31/2005 - 12/29/2006	24	144	12	0	-
BAC	01/03/2005 - 12/01/2006	100	630	25	177	05/23/2008	01/31/2005 - 12/29/2006	24	144	12	26	02/28/2007
BTC	12/07/2014 - 11/27/2016	104	114	10	0	-	12/31/2014 - 11/30/2016	24	26	5	25	12/31/2016
C	01/03/2005 - 12/01/2006	100	630	25	123	05/11/2007	01/31/2005 - 12/29/2006	24	144	12	25	01/31/2007
CSCO	01/03/1998 - 11/27/1999	104	105	10	0	-	01/31/1997 - 12/31/1998	24	24	4	35	11/30/2000
CLX	01/03/2005 - 12/01/2006	100	630	25	0	-	01/31/2005 - 12/29/2006	24	144	12	0	-
DJIA	-	-	-	-	-	-	12/1/1927 - 12/1/1928	12	60	7	17	04/1/1929
DLTR	01/03/2005 - 12/01/2006	100	630	25	0	-	01/31/2005 - 12/29/2006	24	144	12	0	-
GE	01/03/2005 - 12/01/2006	100	630	25	144	05/23/208	01/31/2005 - 12/29/2006	24	144	12	30	06/29/2007
GPRO	01/03/2014 - 06/13/2014	24	212	14	25	12/12/2014	01/31/2014 - 12/31/2014	12	43	6	13	06/30/2015
GS	01/03/2005 - 12/01/2006	100	630	25	178	05/30/2008	01/31/2005 - 12/29/2006	24	144	12	26	02/28/2007
JNJ	01/03/2005 - 12/01/2006	100	630	25	0	-	01/31/2005 - 12/29/2006	24	144	12	0	-
JPM	01/03/2005 - 12/01/2006	100	630	25	0	-	01/31/2005 - 12/29/2006	24	144	12	26	02/28/2007
MSFT	01/03/1998 - 11/27/1999	100	109	10	0	-	01/31/1997 - 12/31/1998	24	24	4	25	01/31/2000
N225	01/08/1988 - 12/28/1988	52	209	14	0	-	01/29/1988 - 12/30/1988	12	48	6	26	02/28/1990
QCOM	01/03/1997 - 11/27/1998	104	105	10	0	-	01/31/1997 - 12/31/1998	24	24	4	25	01/31/2000
WMT	01/03/2005 - 12/01/2006	100	630	25	0	-	01/31/2005 - 12/29/2006	24	144	12	0	-
ZION	01/03/2005 - 12/01/2006	100	630	25	123	05/11/2007	01/31/2005 - 12/29/2006	24	144	12	25	01/31/2007

Table A.2. Actual change dates

Abbreviation	Change Month, Year
AIG, AXP, BAC, C, GE, GS, JPM, ZION	September 2008
BTC	November 2017
CSCO	March 2000
GPRO	October 2016
N225	July 1990
MSFT	March 2000
QCOM	March 2000

Table A.3. Legend for weekly and monthly return plots

Stop Time

Window

Return

For better visualization, if there is no stop time the red line stays at zero and the window is extended from that.

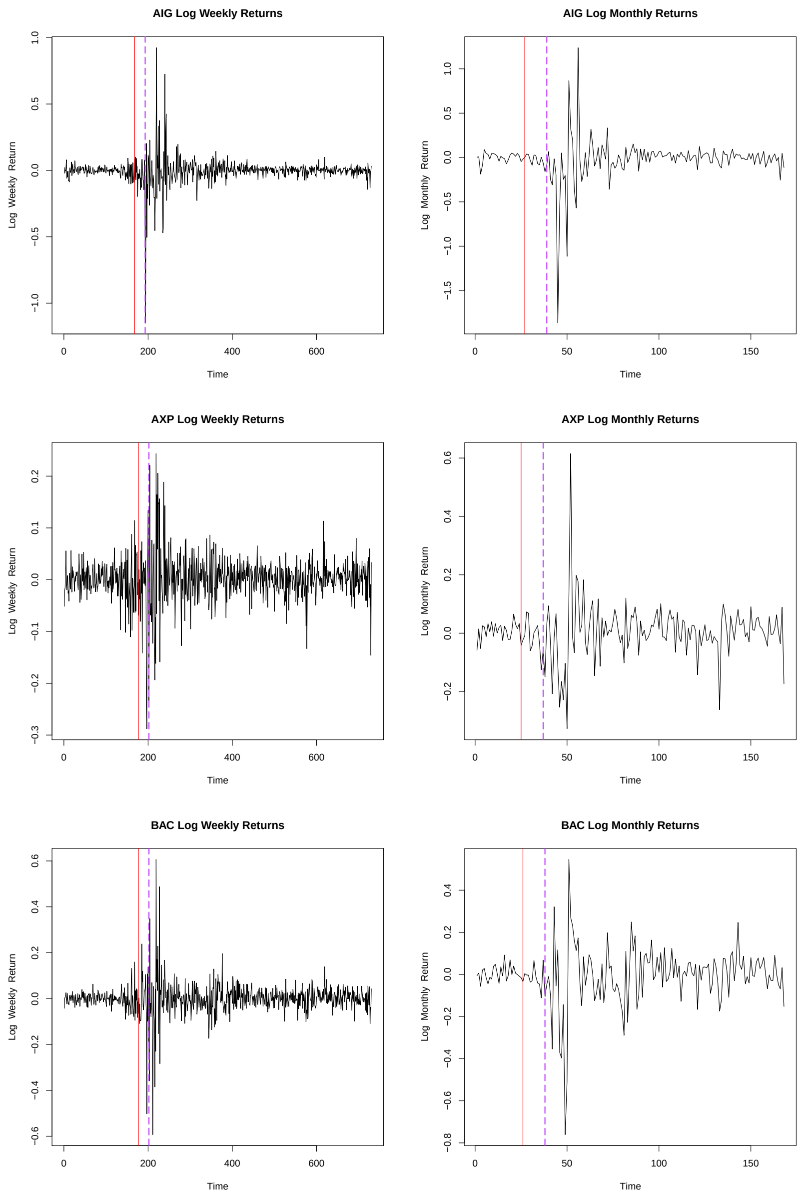


FIGURE A.1. Return stop times for loss group

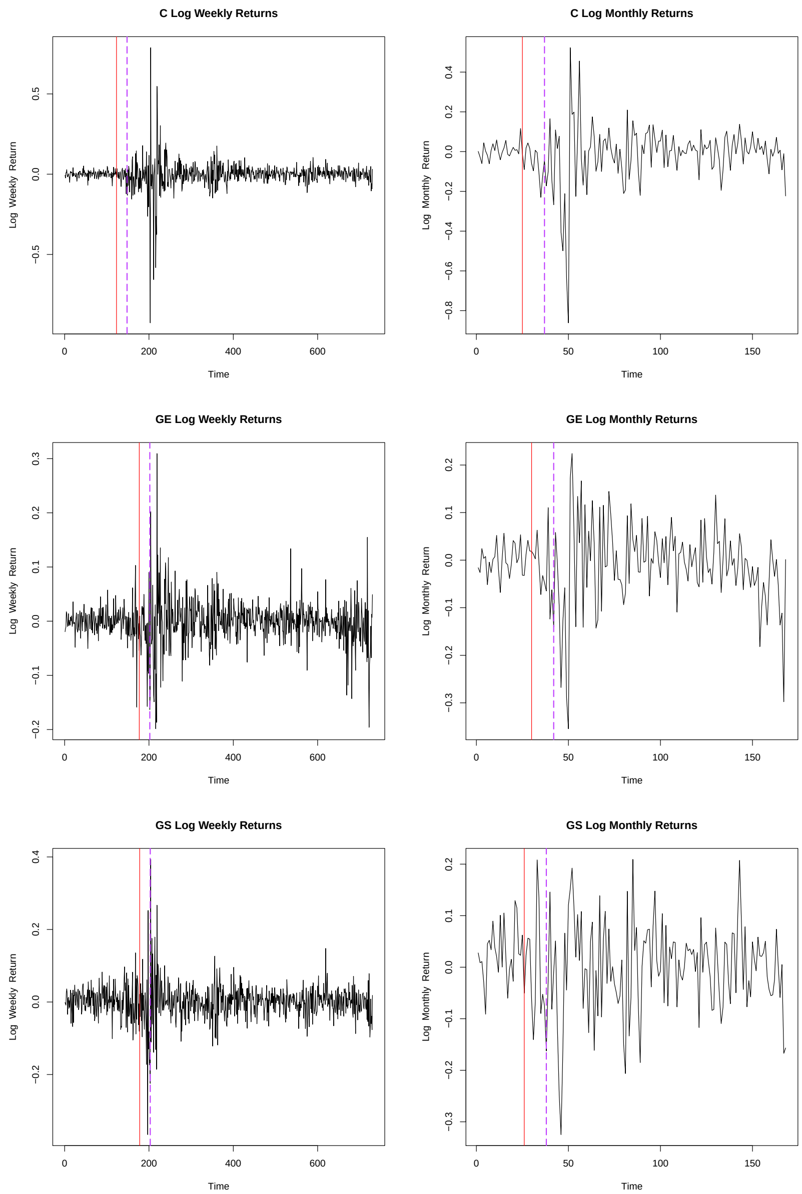


FIGURE A.2. Return stop times for loss group

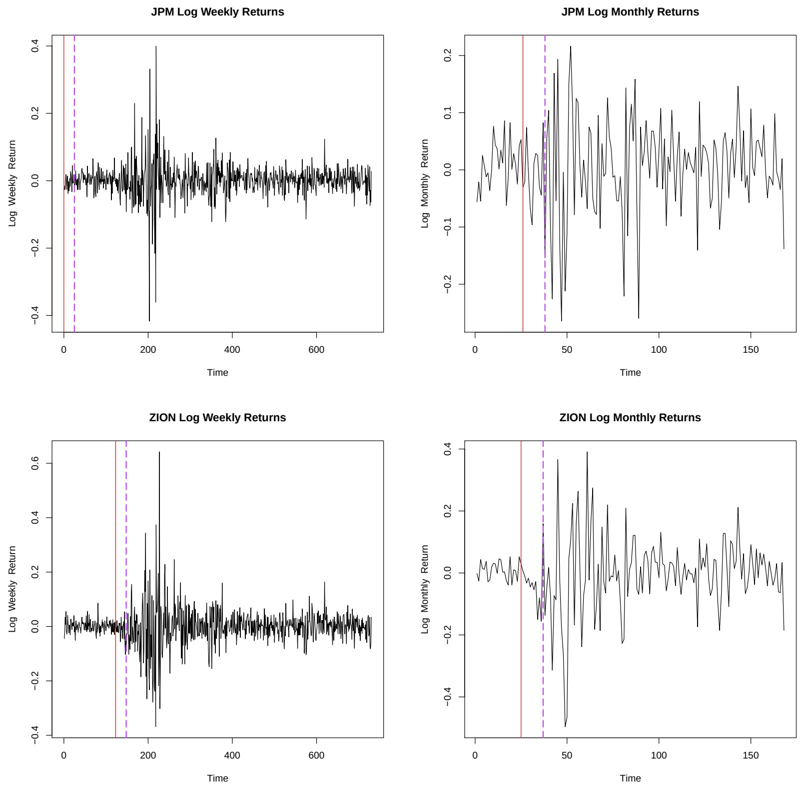


FIGURE A.3. Return stop times for loss group cont.

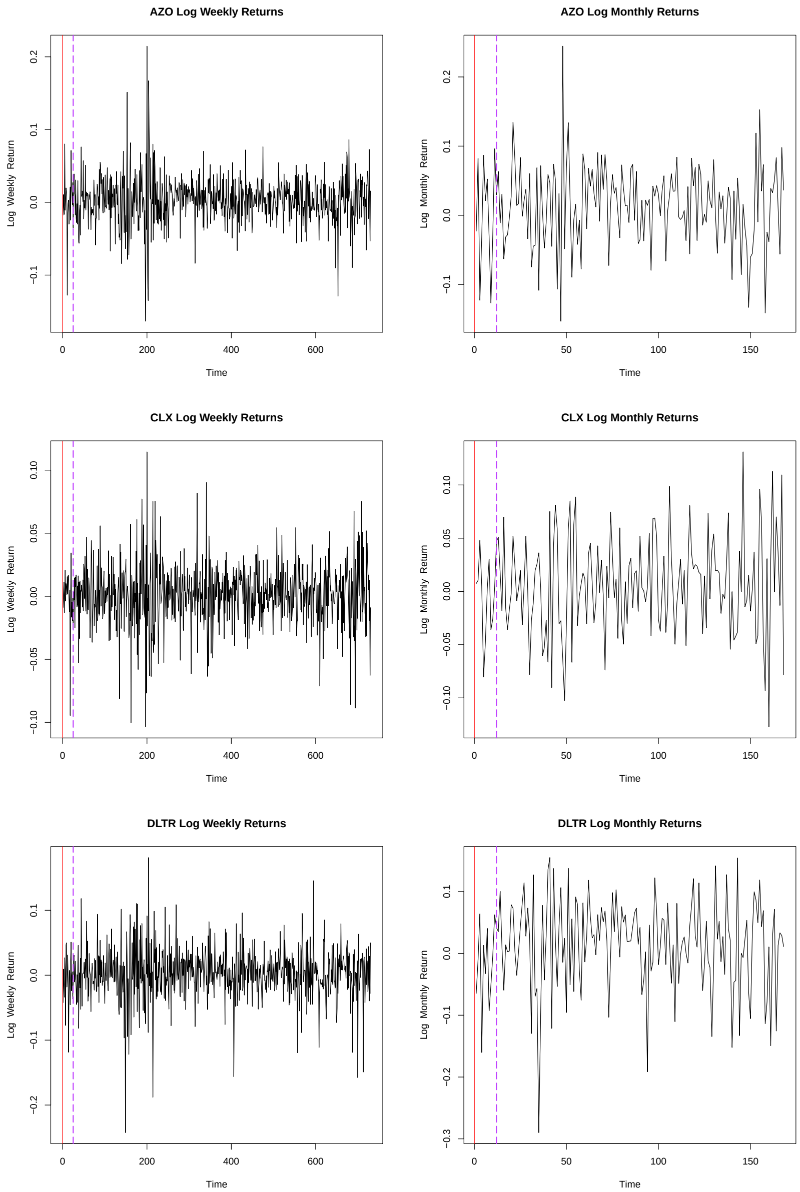


FIGURE A.4. Return stop times for neutral group

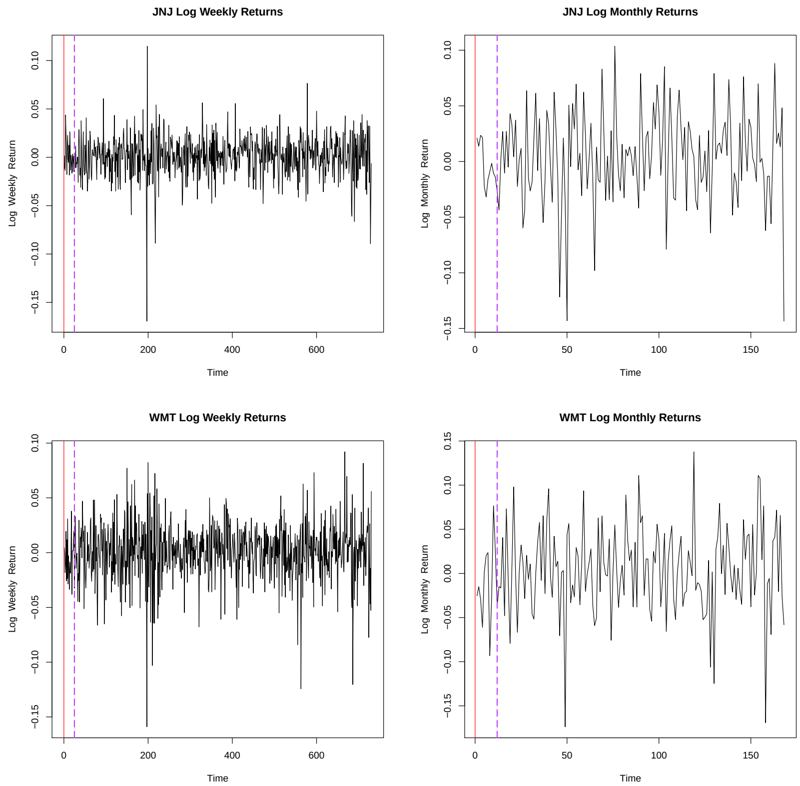


FIGURE A.5. Return stop times for neutral group cont.

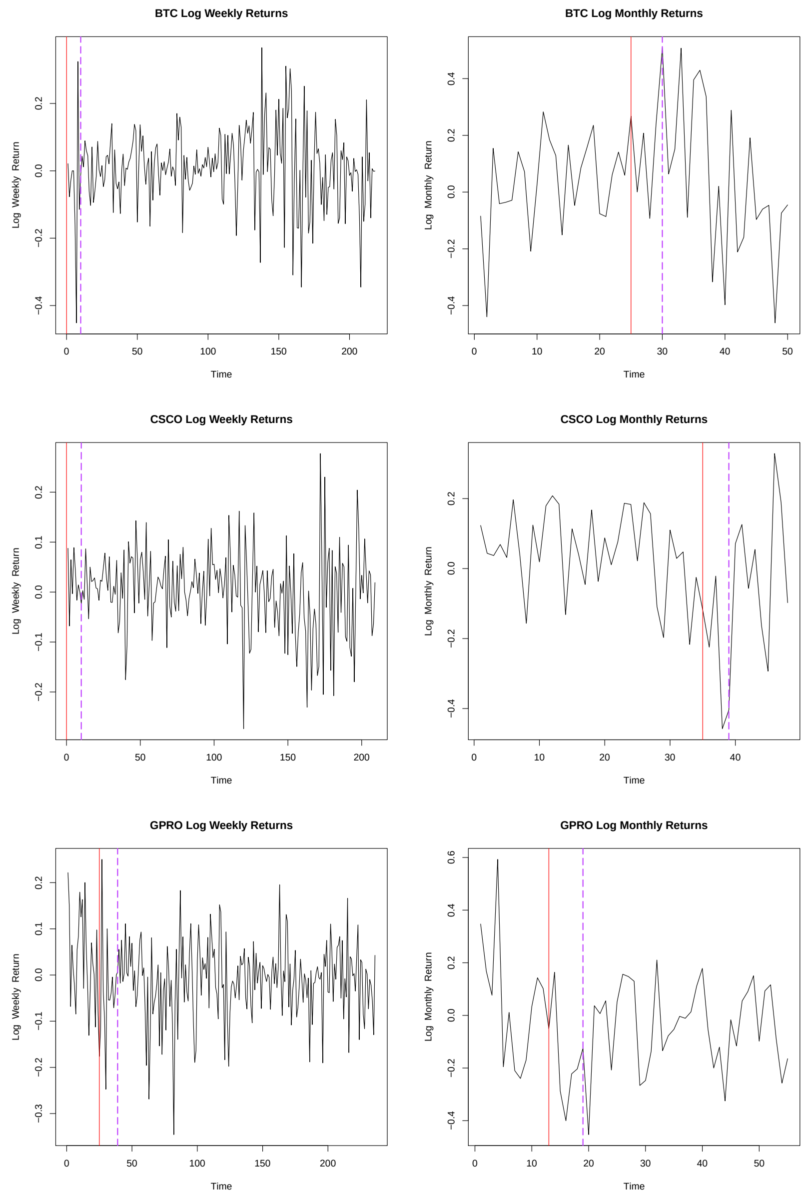


Figure A.6. Return stop times for gain group

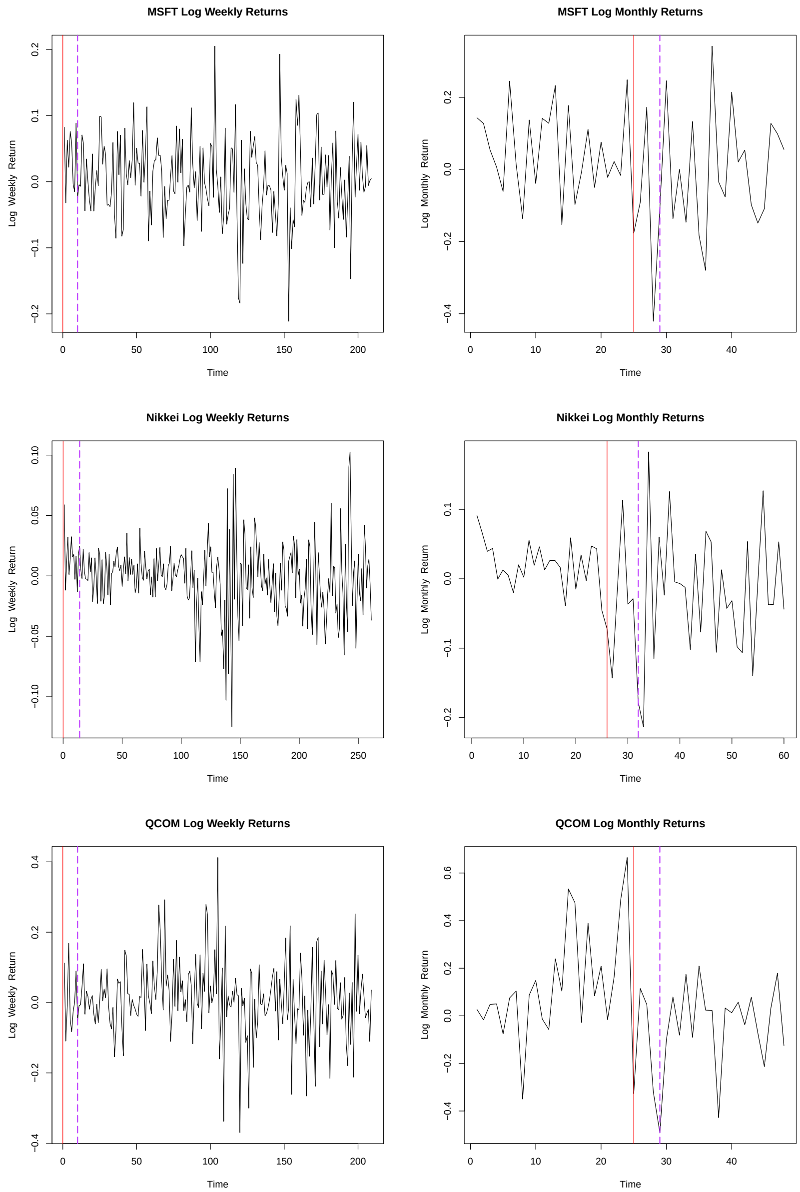


FIGURE A.7. Return stop times for gain group

Appendix B. Proofs

Proof of Theorem 2.1

Proof. First note that rewriting these partial sums as standard Wiener processes assumes that the random variables are mean zero. However, in the context of the test statistic, it actually doesn't matter if we include the mean in the analysis or not. To show this, let $E[X_i] = \mu$. Under the null, this is the mean for all $X_i, i \geq 0$. To get the mean zero random variables in the partial sums we can subtract the mean inside each partial sum

$$\left| \frac{\sum_{i=1}^{M} (X_i - \mu)}{M} - \frac{\sum_{j=M+k}^{M+k+h} (X_j - \mu)}{h} \right| = \left| \frac{\sum_{i=1}^{M} X_i - M\mu}{M} - \frac{\sum_{j=M+k}^{M+k+h} X_j - h\mu}{h} \right|$$

$$= \left| \frac{\sum_{i=1}^{M} X_i}{M} - \mu - \left(\frac{\sum_{j=M+k}^{M+k+h} X_j}{h} - \mu \right) \right|$$

$$= \left| \frac{\sum_{i=1}^{M} X_i}{M} - \frac{\sum_{j=M+k}^{M+k+h} X_j}{h} \right|.$$

We use this result and Assumption 2.1 we can assume that $\sigma = 1$ we can rewrite the test statistic in the following way

$$\max_{1 \leq k \leq T-h} \left| \frac{W_1(M)}{M} - \frac{1}{h}(W_2(h)) \right| \Big/ g(h,k)$$

$$= \max_{1 \leq k \leq T-h} \left| \frac{W_1(M)}{M} - \frac{1}{h}(W_2(k+h) - W_2(k)) \right| \Big/ g(h,k)$$

$$= \max_{1 \leq k \leq T-h} \left| \frac{W_1(M)}{\sqrt{M}} \frac{1}{\sqrt{M}} - \frac{1}{\sqrt{h}} \frac{(W_2(k+h) - W_2(k))}{\sqrt{h}} \right| \Big/ g(h,k)$$

$$= \max_{1 \leq k \leq T-h} \left| \frac{W_1(M)}{\sqrt{M}} \frac{1}{\sqrt{M}} - \frac{1}{\sqrt{h}} \frac{(W_2(k+h) - W_2(k))}{\sqrt{h}} \right| \Big/ g(h,k)$$

$$= \max_{1 \leq k \leq T-h} \left| \frac{\sqrt{h}}{\sqrt{M}} \frac{W_1(M)}{\sqrt{M}} - \frac{(W_2(k+h) - W_2(k))}{\sqrt{h}} \right| \Big/ \left(\sqrt{h} g(h,k) \right).$$

Assume that $h/M \to 0$ and this implies that $\sqrt{h/M} \to 0$. Then we can simplify the expression above to be

$$\begin{aligned}
&\max_{1\le k\le T-h} \frac{1}{\sqrt{h}} \left| \frac{W_2(k+h)-W_2(k)}{\sqrt{h}} \right| \Big/ g(h,k) \\
&= \max_{1\le k\le T-h} \frac{|W_2(k+h)-W_2(k)|}{\sqrt{h}} \Big/ \left(\sqrt{h}g(h,k)\right)
\end{aligned}$$

Now let $k = uh$ for some $u \in \mathbb{R}$. For $\beta > 1/2$ with the law of iterated logarithm

$$\max_{0<u<\infty} |W_2(u+1)-W_2(u)| \frac{1}{(u+1)^\beta}$$

is well defined. So we want to have $\sqrt{h}g(h,uh) = (u+1)^\beta$

$$\begin{aligned}
&\max_{1\le k\le T-h} \frac{|W_2(k+h)-W_2(k)|}{\sqrt{h}} \Big/ \left(\sqrt{h}g(h,k)\right) \\
&= \max_{1\le uh\le T-h} \frac{|W_2(uh+h)-W_2(uh)|}{\sqrt{h}} \Big/ \left(\sqrt{h}g(h,uh)\right) \\
&= \max_{1\le uh\le T-h} \frac{|W_2(uh+h)-W_2(uh)|}{\sqrt{h}} \Big/ \left(\sqrt{h}g(h,uh)\right) + o_{\mathrm{P}}(1),
\end{aligned}$$

which is close in distribution to

$$\max_{1/h\le u\le (T-h)/h} |W_2(u+1)-W_2(u)| \Big/ \left(\sqrt{h}g(h,uh)\right).$$

To derive the desired function, we need that $1/h \to 0$ and $T/h \to \infty$, for this we need that $h \to \infty$ but also that $\sqrt{h}(g(h,uh)) \to (1+u)^\beta$ and the limit is

$$\max_{0\le u\le\infty} |W_2(u+1)-W_2(u)| /(1+u)^\beta, \tag{B.1}$$

since

$$\sqrt{h}g(h,k) = \sqrt{h}g(h,uh) = (u+1)^\beta = \frac{\sqrt{h}h^\beta}{\sqrt{h}h^\beta}(u+1)^\beta = \sqrt{h}\frac{(h+uh)^\beta}{h^{\beta+1/2}}$$

and this gives us our boundary function

$$g(h,k) = \frac{(h+k)^\beta}{h^{\beta+1/2}}.$$

Now assume that the variance for all X_i is σ^2, we can follow a similar derivation

$$\max_{1\le k\le T-h}\left|\frac{\sigma W_1(M)}{M}-\frac{1}{h}\left(\sigma W_2(h)\right)\right| \bigg/ g(h,k)$$

$$= \max_{1\le k\le T-h}\left|\frac{\sigma W_1(M)}{M}-\frac{1}{h}\left(\sigma W_2(k+h)-\sigma W_2(k)\right)\right| \bigg/ g(h,k)$$

$$= \max_{1\le k\le T-h}\left|\frac{\sigma W_1(M)}{\sqrt{M}}\frac{1}{\sqrt{M}}-\frac{1}{\sqrt{h}}\frac{\sigma(W_2(k+h)-W_2(k))}{\sqrt{h}}\right| \bigg/ g(h,k)$$

$$= \max_{1\le k\le T-h}|\sigma|\left|\frac{W_1(M)}{\sqrt{M}}\frac{1}{\sqrt{M}}-\frac{1}{\sqrt{h}}\frac{(W_2(k+h)-W_2(k))}{\sqrt{h}}\right| \bigg/ g(h,k)$$

$$= \max_{1\le k\le T-h}|\sigma|\left|\frac{\sqrt{h}}{\sqrt{M}}\frac{W_1(M)}{\sqrt{M}}-\frac{(W_2(k+h)-W_2(k))}{\sqrt{h}}\right| \bigg/ \left(\sqrt{h}g(h,k)\right).$$

Assume that $h/M \to 0$ and this implies that $\sqrt{h/M} \to 0$. Then we can simplify the expression above to be

$$= \max_{1\le k\le T-h}\frac{|\sigma|}{\sqrt{h}}\left|\frac{W_2(k+h)-W_2(k)}{\sqrt{h}}\right| \bigg/ g(h,k)$$

$$= \max_{1\le k\le T-h}\frac{|\sigma|}{\sqrt{h}}\frac{|W_2(k+h)-W_2(k)|}{\sqrt{h}} \bigg/ g(h,k)$$

$$= \max_{1\le k\le T-h}|\sigma|\,\frac{|W_2(k+h)-W_2(k)|}{\sqrt{h}} \bigg/ \left(\sqrt{h}g(h,k)\right)$$

Now let $k = uh$ for some $u \in \mathbb{R}$. For $\beta > 1/2$ with the law of iterated logarithm

$$= \max_{0<u<\infty}|(W_2(u+1)-W_2(u))|\,\frac{1}{(u+1)^\beta}. \tag{B.2}$$

We want to have

$$\left(\frac{\sqrt{h}}{\sigma}g(h,k)) = (k+h)^\beta\right).$$

So we can rewrite (B.2) as

$$\max_{1/h\le u\le (T-h)/h}|\sigma|\,\frac{(W_2(h(u+1))-W_2(h(u)|}{\sqrt{h}} \bigg/ \left(\sqrt{h}g(h,uh)\right)$$

$$\overset{\mathcal{D}}{=}\frac{1}{h^\beta}\max_{1/h\le u\le (T-h)/h}|\sigma|\,|(W_2(u+1)-W_2(u))| \bigg/ (1+u)^\beta.$$

We need that

$$\frac{\sqrt{h}}{|\sigma|}g(h,k) = \frac{\sqrt{h}}{|\sigma|}g(h,uh) = (u+1)^{\beta} = \frac{\sqrt{h}\,|\sigma|}{\sqrt{h}\,|\sigma|}\frac{h^{\beta}(u+1)^{\beta}}{h^{\beta}}$$

$$= \frac{\sqrt{h}\,|\sigma|}{\sqrt{h}\,|\sigma|}\frac{(h+hu)^{\beta}}{h^{\beta}} = \frac{\sqrt{h}\,|\sigma|}{\sqrt{h}\,|\sigma|}\frac{(h+k)^{\beta}}{h^{\beta}}$$

and this gives us our boundary function

$$g(h,k) = \frac{|\sigma|\,(h+k)^{\beta}}{\sqrt{h}h^{\beta}} = \frac{|\sigma|\,(h+k)^{\beta}}{h^{\beta+1/2}}. \tag{B.3}$$

□

Proof of Theorem 2.2

Proof. We start by rewriting the probability of the sequential monitoring procedure finding a change point,

$$\mathrm{P}(\tau_M < T-h) = \mathrm{P}\left(\frac{Z_k}{g_{\alpha}(h,k)} > 1 \text{ for } 1 \le k \le T-h\right) = \mathrm{P}\left(\max_{1\le k\le T-h}\frac{Z_k}{g(h,k)} \le c_{\alpha}\right).$$

According to our calculations

$$\max_{1\le k\le T-h}\frac{|Z_k|h^{\beta+1/2}}{\sigma(h+k)^{\beta}} = \max_{1\le k\le T-h}\frac{\left|\bar{X}_M - \frac{1}{h}\sum\limits_{i=M+k+1}^{M+k+h}X_i\right| h^{1/2+\beta}}{\sigma(h+k)^{\beta}} \tag{B.4}$$

$$\max_{1\le k\le T-h}\frac{\left|\bar{X}_M\right| h^{1/2+\beta}}{\sigma(h+k)^{\beta}} = o_{\mathrm{P}}\left(\frac{1}{\sqrt{M})}\frac{h^{1/2+\beta}}{h^{\beta}}\right) = o_{\mathrm{P}}\left(\left(\frac{h}{M}\right)^{1/2}\right) = o_{\mathrm{P}}(1), \tag{B.5}$$

and therefore

$$\begin{aligned}
\max_{1\le k\le T-h}\frac{|Z_k|h^{\beta+1/2}}{\sigma(h+k)^{\beta}} &\approx \max_{1\le k\le T-h}\frac{\left|\frac{1}{h}\sum\limits_{i=M+k+1}^{M+k+h}X_i\right| h^{1/2+\beta}}{\sigma(h+k)^{\beta}} + o_{\mathrm{P}}(1) \\
&= \max_{1\le k\le T-h}\frac{1}{h}\frac{\left|\sum\limits_{i=M+k+1}^{M+k+h}X_i\right| h^{1/2+\beta}}{\sigma(h+k)^{\beta}} \\
&\approx \max_{1\le k\le T-h}\frac{h^{\beta-1/2}(M+k+h)^{\alpha}}{(h+k)^{\beta}} \quad \text{for } \alpha < 1/2
\end{aligned}$$

It is necessary here to check how this quantity behaves for the extreme values of k, first let $k = 1$

$$\frac{h^{\beta+1/2}(M+k)^{\alpha}}{h^{\beta}} = \frac{1}{\sqrt{h}}(M+h)^{\alpha} \approx \frac{M^{\alpha}}{\sqrt{h}} \to 0 \text{ as } h \to \infty. \tag{B.6}$$

Now let $k = T$

$$\begin{aligned}\frac{h^{\beta+1/2}(M+T+h)^{\alpha}}{(h+T)^{\beta}} &\approx \frac{h^{1/2}(M+T)^{\alpha}}{T^{\beta}} \\ &\leq \frac{(M^{\alpha}+T^{\alpha})}{\sqrt{h}T^{\beta}} \to \frac{1}{T^{\beta-\alpha}\sqrt{h}} \to 0,\end{aligned}$$

thus we have

$$\max_{1\leq k\leq T-h} \frac{\frac{1}{h}\left|W(M+k+h)-W(M+k+1)\right| h^{1/2+\beta}}{(h+k)^{\beta}} \overset{\mathcal{D}}{=} \max_{1\leq k\leq T-h} \frac{\left|W(k+h)-W(k+1)\right| h^{1/2+\beta}}{(h+k)^{\beta}}.$$

Let $k = uh$, we assume that $h \to \infty$ and that $T/h \to \infty$ to rewrite the previous quantity as

$$\begin{aligned}&\max_{1\leq uh\leq T-h} \frac{\left|W(k+h)-W(k+1)\right| h^{1/2+\beta}}{(h+k)^{\beta}} \\ &= \max_{1/h\leq u\leq (T-h)/h} \frac{\left|W(u+1)-W(u+1/h)\right| h^{1/2+\beta}}{(u+1)^{\beta}} \xrightarrow{\mathcal{D}} \max_{0<u<\infty} \frac{\left|W(u+1)-W(u)\right|}{(u+1)^{\beta}}\end{aligned}$$

Here we can choose c to be the critical values that were simulated in Section 4. $\square$

Proof of Theorem 3.3

Proof. We divide the proof into steps.

Step 1: For integers $k \in \mathbb{Z}$, by Feller's inequality for the named distribution we have that for all $\varepsilon > 0$ there is $0 < \delta < 1$ such that by the Borel - Cantelli lemma

$$\mathrm{P}\left(|\Gamma(k)| \geq (1-\varepsilon)\sqrt{2\log(k)}\right) = 2\left(1 - \Phi\left((1-\varepsilon)\sqrt{2\log(k)}\right)\right) \geq k^{1-\delta}.$$

Hence

$$\sum_{k=1}^{\infty} \mathrm{P}\left(|\Gamma(y)| \geq (1-\varepsilon)\sqrt{2\log(k)}\right) \geq \sum_{k=1}^{\infty} k^{1-\delta} = \infty \text{ for all } \varepsilon > 0.$$

Since $\{\Gamma(k), k \geq 1\}$ are independent random variables,

$$1 - \varepsilon \leq \liminf_{k\to\infty} \frac{|\Gamma(k)|}{\sqrt{2\log k}} \text{ a.s.}$$

Since $\varepsilon > 0$ is arbitrary $\varepsilon \to 0$

$$1 \leq \limsup_{k\to\infty} \frac{|\Gamma(k)|}{\sqrt{2\log k}} \quad \text{a.s.}$$

Let $t_k = ck$ where $c \in \mathbb{R}$ we want to show that

$$\limsup_{k\to\infty} \frac{|\Gamma(t_k)|}{\sqrt{2\log t_k}} \leq 1 \quad \text{a.s.} \tag{B.7}$$

Using that $\Gamma(t_k) \stackrel{\mathcal{D}}{=} N(0,1)$ we get again

$$\mathrm{P}\left\{|\Gamma(t_k)| \geq (1+\varepsilon)\sqrt{2\log(t_k)}\right\} = 2\left(1 - \Phi\left((1+\varepsilon)\sqrt{2\log(t_k)}\right)\right)$$

Using the well known Feller's Inequality (cf. Feller (1968), p. 175) we can give an upper bound to the latter.

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\frac{e^{x^2/2}}{\sqrt{2\pi}} \leq \mathrm{P}(X \geq x) \leq \frac{1}{x}\frac{e^{x^2/2}}{\sqrt{2\pi}} \quad \text{for all } x > 0. \tag{B.8}$$

Using the upper bound of (B.8) we can form a convergent series.

$$\begin{aligned}
2\left(1 - \Phi\left((1+\varepsilon)\sqrt{2\log(t_k)}\right)\right) &\leq \frac{ce^{-((1+\varepsilon)\sqrt{2\log(t_k)})^2/2}}{(1+\varepsilon)\sqrt{2\log(t_k)}} \\
&= \frac{ce^{-(1+\varepsilon)^2\log(t_k)}}{(1+\varepsilon)\sqrt{2\log(t_k)}} \\
&= \frac{c}{1+\varepsilon}\frac{1}{\sqrt{2\log(t_k)}}t_k^{-(1+\varepsilon)^2} = \frac{c}{(1+\varepsilon)\sqrt{2}}c^{-(1+\varepsilon)^2}\frac{k^{-(1+\varepsilon)^2}}{\log(ck)}
\end{aligned}$$

This is a convergent series. Thus we have,

$$\sum_{k=1}^{\infty} \mathrm{P}\left(|\Gamma(t_k)| \geq (1+\varepsilon)\sqrt{2\log(t_k)}\right) < \infty$$

for any $c \in \mathbb{R}$. By the Borel-Cantelli lemma we have B.7. Now we combine the previous two into the following quantity in the form of

$$\sup_{t_k \leq y \leq t_{k+1}} |W(y+1) - W(y) - (W(t_k+1) - W(t_k))|.$$

By the triangle inequality this is bounded above

$$\begin{aligned}
&\sup_{t_k \leq y \leq t_{k+1}} |W(y+1) - W(y) - (W(t_{k+1}) - W(t_k))| \\
&\leq \sup_{t_k \leq y \leq t_{k+1}} |W(y+1) - W(t_{k+1})| + \sup_{t_k \leq y \leq t_{k+1}} |W(y) - W(t_k)|
\end{aligned}$$

By the properties of $W(t)$

$$\sup_{t_k \le y \le t_{k+1}} |W(y) - W(t_k)| = \sup_{0 \le x \le t_{k+1} - t_k} |W(x)|$$
$$= \sup_{0 \le x \le c} |W(x)| \stackrel{\mathcal{D}}{=} c^{1/2} \sup_{0 \le x \le 1} |W(x)|,$$

and therefore

$$\mathrm{P}\left(c^{1/2} \sup_{t_k \le y \le t_{k+1}} |W(y) - W(t_k)| \ge (1+\varepsilon)\sqrt{2\log(t_{k+1})} \right)$$

$$= \mathrm{P}\left(\sup_{0 \le x \le 1} |W(x)| \ge (1+\varepsilon)\sqrt{2\log(t_{k+1})} \right) \le \frac{c\delta}{(1+\varepsilon)\sqrt{2\log(t_{k+1})}} e^{\frac{-(1-\delta)(1+\varepsilon)^2 2\log(t_{k+1})}{2}}.$$

Hence

$$\sum_{k=1}^{\infty} \mathrm{P}\left(\sup_{0 \le x \le 1} |W(x)| \ge (1+\varepsilon)\sqrt{2\log(t_{k+1})} \right) \le \sum_{k=1}^{\infty} \frac{c\delta}{(1+\varepsilon)\sqrt{2\log(t_{k+1})}} e^{\frac{-(1-\delta)(1+\varepsilon)^2 2\log(t_{k+1})}{2}} < \infty$$

Combining this with statement 2 and applying Borel-Cantelli

$$\mathrm{P}\left(\limsup_{k\to\infty} c^{-1/2} \frac{|\Gamma(t_k) - \Gamma(y)|}{\sqrt{2\log(t_{k+1})}} \ge (1+\varepsilon) \right) = 0.$$

Thus

$$\limsup_{k\to\infty} c^{-1/2} \frac{|\Gamma(t_k) - \Gamma(y)|}{\sqrt{2\log(t_{k+1})}} \le 1 \text{ a.s.}, \qquad \limsup_{k\to\infty} \frac{|\Gamma(t_k) - \Gamma(y)|}{\sqrt{2\log(t_{k+1})}} \le c^{1/2} \text{ a.s.}$$

Combine this with the first statement to get that

$$\limsup_{y\to\infty} \frac{|\Gamma(y)|}{\sqrt{2\log(y)}} \le 1 + c^{1/2} \text{ a.s.}$$

For any $c \in \mathbb{R}$ so choose $c = 1$

$$\limsup_{y\to\infty} \frac{|\Gamma(y)|}{\sqrt{2\log(y)}} = 1 \text{ a.s.}$$

$\square$

Appendix C. Reference for Generalized Autoregressive (GARCH) Models

It is often the case with financial and economic time series data that the volatility tends to be autocorrelated. One model to capture this autocorrelation is the ARCH process due to Engle (1982). To illustrate, let r_t denote the daily return on an asset and assume that $\mathrm{E}[r_t] = 0$. An ARCH(1) model for r_t is

$$
\begin{aligned}
r_t &= \sigma_t z_t \\
z_t &\sim \text{i.i.d. } N(0,1) \\
\sigma_t^2 &= \omega + \alpha r_{t-1}^2
\end{aligned}
\tag{C.1}
$$

where $\omega > 0$ and $0 < \alpha_1 < 1$. It is important to note that r_t is an uncorrelated process based on this formulation. That is, if we let $I_t = \{r_t, r_t - 1, ...\}$ be the conditioning information set of returns up to time t we can take the conditional expectation

$$
\begin{aligned}
\mathrm{E}[r_t|I_t] &= \mathrm{E}[\sigma_t z_t|I_t] \\
&= \sigma_t \mathrm{E}[z_t|I_t] \\
&= \sigma_t \mathrm{E}[z_t] = 0,
\end{aligned}
\tag{C.2}
$$

provided that $\alpha < 1$, r_t is a mean zero, covariance stationary process with a finite variance. The unconditional variance of r_t is given by

$$
\begin{aligned}
\mathrm{Var}(r_t) &= \mathrm{E}[r_t^2] = \mathrm{E}[\mathrm{E}[z_t^2\sigma_t^2]] \\
&= \mathrm{E}[\sigma_t^2 \mathrm{E}[z_t^2|I_{t-1}] = \mathrm{E}[\sigma_t^2]
\end{aligned}
$$

Utilizing (C.1) and the stationarity of $r_t, \mathrm{E}[\sigma_t^2]$ may be expressed as

$$
\mathrm{E}[\sigma_t^2] = \frac{\omega}{1-\alpha_1} \tag{C.3}
$$

If r_t is stationary and ergodic then again we have that the sample mean

$$
\bar{r} \overset{\mathcal{D}}{=} N\left(\mu, \frac{1}{T}\frac{\omega}{1-\alpha_1}\right)
$$

So the long-run variance is indeed

$$
\sigma_{r_t}^2 = \frac{\omega}{1-\alpha_1}. \tag{C.4}
$$

We obtain an ARCH(p) process if r_t^2 follows an AR(p) process. And we can write

$$
\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2.
$$

Similar to C.4, utilizing again the structure of r_t and r_t^2 we have that

$$
\sigma_{r_t}^2 = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i}. \tag{C.5}
$$

The stationarity condition holds if

$$0 < \alpha_1 + ... + \alpha_p < 1. \tag{C.6}$$

The generalized ARCH or GARCH model is an alternative to the ARCH(p) model. It is given by

$$\sigma_t^2 = \omega + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

where the ARCH term is r_{t-1}^2 and the GARCH term is σ_{t-1}^2. GARCH(p, q) models include p ARCH terms and q GARCH terms. The unconditional variance for a GARCH(1,1) process is

$$\mathrm{Var}(r_t) = \sigma_{r_t}^2 = \frac{\omega}{1 - \alpha_1 - \beta_1}. \tag{C.7}$$

Similar to the ARCH process, this process is stationary if the following condition holds

$$0 < \alpha_1 + \beta_1 < 1. \tag{C.8}$$

For GARCH(p, q) we write σ_t^2 as

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$$

Using a similar derivation as the previous models, we have that the unconditional variance of this process is

$$\sigma_{r_t}^2 = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}.$$

.

Proof. For GARCH(p, q), we have the following setup

$$\begin{aligned} r_t &= \sigma_t z_t \\ z_t &\sim \text{i.i.d. } N(0,1) \\ \sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \end{aligned}$$

We want the unconditional variance, σ^2 of the process r_t. First recall that r_t is a mean zero, uncorrelated process

$$\begin{aligned} \mathrm{E}[r_t] &= 0 \\ \mathrm{Cov}(r_i, r_j) &= 0 \ \text{ for } i \neq j. \end{aligned}$$

We need to isolate r_t^2 to one side, we rewrite the process as

$$
\begin{aligned}
r_t^2 - \sigma_t^2 &= \sigma_t^2(\varepsilon_t^2 - 1) \\
\sum_{j=1}^{q} \beta_j (r_{t-j}^2 - \sigma_{t-j}^2) &= \sum_{j=1}^{q} \beta_j \sigma_{t-j}(\varepsilon_{t-j} - 1)
\end{aligned}
$$

where we let $\nu_t = \sigma_t^2(\varepsilon_t^2 - 1)$ which has mean zero because ε_t^2 is a $\chi^2_{(1)}$ random variable. We then subtract the above equations to get

$$
\begin{aligned}
r_t^2 - \sigma_t^2 - \sum_{j=1}^{q} \beta_j (r_{t-j}^2 - \sigma_{t-j}^2) &= \nu_t - \sum_{j=1}^{q} \beta_j \nu_{t-j} \\
r_t^2 - \sigma_t^2 - \sum_{j=1}^{q} \beta_j r_{t-j}^2 - \beta_j \sigma_{t-j}^2 &= \nu_t - \sum_{j=1}^{q} \beta_j \nu_{t-j} \\
r_t^2 - \sum_{j=1}^{q} \beta_j r_{t-j}^2 - (\sigma_t^2 - \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2) &= \nu_t - \sum_{j=1}^{q} \beta_j \nu_{t-j} \\
r_t^2 - \sum_{j=1}^{q} \beta_j r_{t-j}^2 - (\omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2) &= \nu_t - \sum_{j=1}^{q} \beta_j \nu_{t-j} \\
r_t^2 = \omega + \sum_{j=1}^{q} \beta_j r_{t-j}^2 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \nu_t - \sum_{j=1}^{q} \beta_j \nu_{t-j}.
\end{aligned}
$$

Now note that $\mathrm{E}[r_t^2] - \mathrm{Var}(r_t^2) = \sigma^2$ because r_t has mean zero. So by taking expectation we obtain

$$
\begin{aligned}
\sigma^2 = \mathrm{E}[r_t^2] &= \omega + \sum_{j=1}^{q} \beta_j \mathrm{E}[r_{t-j}^2] + \sum_{i=1}^{p} \alpha_i \mathrm{E}[r_{t-i}^2] + \mathrm{E}[\nu_t] + \sum_{j=1}^{q} \beta_j \mathrm{E}[\nu_{t-j}] \\
&= \omega + \sum_{j=1}^{q} \beta_j \sigma^2 + \sum_{i=1}^{p} \alpha_i \sigma^2.
\end{aligned}
$$

Therefore,

$$
\sigma^2 = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}.
$$

$\square$

References

[1] Adrian Trapletti and Kurt Hornik (2018). tseries: Time Series Analysis and Computational Finance. R package version 0.10-46.

[2] Aue, A. and Horváth, L. Structural breaks in time series, *Journal of Time Series Analysis*, 2012.

[3] Benzinga, W. B. The best and worst stocks of the 2008 crash: what we learned, `https://finance.yahoo.com/news/best-worst-stocks-2008-crash-205844396.html` (updated April 14, 2016).

[4] Bown, J. Change Point Research, (2019), GitHub repository, `https://github.com/GuitarPro6/change_point_research`

[5] Csörgő, M. and Révész, P. *Strong Approximations in Probability and Statistics.* Academic Press, 1981.

[6] DasGupta, A. *Asymptotic Theory of Statistics and Probability*, Springer, 2008.

[7] Debin Qiu (2015). aTSA: Alternative Time Series Analysis. R package version 3.1.2. `https://CRAN.R-project.org/package=aTSA`

[8] Diethelm Wuertz, Tobias Setz, Yohan Chalabi, Chris Boudt, Pierre Chausse and Michal Miklovac (2017). fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling. R package version 3042.83. `https://CRAN.R-project.org/package=fGarch`

[9] Engle, R.F. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflations, Econometrica, 1982.

[10] Federal Reserve Bank of St. Louis and US. Office of Management and Budget, Dow-Jones Industrial Stock Price Index for United States[M1109BUSM293NNBR], retrieved from FRED, Federal Reserve Bank of St. Louis; `https://fred.stlouisfed.org/series/M1109BUSM293NNBR,January29,2019.`

[11] Fuller, W.A. *Introduction to Statistical Time Series, Second Edition.* John Wiley & Sons, New York, 1996.

[12] Hamilton, J.D. *Time Series Analysis. Princeton University Press* New Jersey, 1994.

[13] Jeffrey A. Ryan and Joshua M. Ulrich (2018). xts: eXtensible Time Series. R package version 0.11-2. `https://CRAN.R-project.org/package=xts`

[14] Jeffrey A. Ryan and Joshua M. Ulrich (2018). quantmod: Quantitative Financial Modeling Framework. R package version 0.4-13. `https://CRAN.R-project.org/package=quantmod`

[15] Jesse Vent (2019). crypto: Cryptocurrency Market Data. R package version 1.1.1. `https://CRAN.R-project.org/package=crypto`

[16] Khoshnevisan, D. *Probability*, 2013.

[17] Newman, C.M. and Wright, A.L. (1981). *An Invariance Principle For Certain Dependent Sequences.*

[18] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/.`

[19] Shumway, R. H. and Stoffer, D. S. *Time Series Analysis and Its Applications With R Examples, Fourth Edition*, Springer, 2017.

[20] Wiener, N. Differential space. J. Math. Phys. 2, 131-174, 1923.

[21] Wang, J., Zivot, E., *Modeling Financial Time Series With S-PLUS®*, Springer, 2006.

[22] Zeileis A (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. Journal of Statistical Software , `http://doi.org/10.18637/jss.v011.i10.`